

Dr Healy, Dr Kennedy and Mr Dingell

The new director of the NIH defended herself before the Dingell committee last week but also promised self-restraint. But she has a strong case for reforming the OSI.

NOBODY disputes that Mr John Dingell, the chairman of the US House of Representatives investigations and oversight sub-committee, is a powerful figure, but two events last week will have confirmed his influence. First, Dr Donald Kennedy resigned as president of Stanford University just six months after the Dingell committee has wrung from him the admission that Stanford's claims for the indirect costs of administering research grants had been inflated by costs whose relevance to the administration of research cannot be imagined. Then, last week, Dr Bernadine Healy, the new director of the National Institutes of Health (NIH) told the Dingell committee that she will not intervene in the administration of NIH's policies on scientific misconduct until a case she oversaw at the Cleveland Clinic has been disposed of.

Each development is a serious humiliation for the research enterprise. Although, in Kennedy's case, it is more likely that the practice of charging illicit spending to the indirect costs budget was cultivated by over-zealous administrators rather than by Kennedy personally, the fact that the Dingell committee rather than Kennedy himself brought the shabby circumstances to light was bound to be damaging. That the indirect costs system may have been widely abused is one thing. That university administrations may sometimes allow their presidents' personal spending to inflate academic costs more generally will strengthen the general but false impression that campus communities are bastions of self-defined privilege. Yet it is a great misfortune that it is Kennedy who is the first victim of Dingell's disclosures: he has been an excellent president at Stanford as, before that, he did outstanding public service at the Food and Drug Administration.

Healy's humiliation last week (see page 461) is more complicated, amounting to a public demonstration by the Dingell committee that the director of NIH is not the master of everything he or she is responsible for. Large parts of NIH, the National Cancer Institute for example, appear as separate lines in the congressional budget, which means that the director of NIH cannot simply take that money and use it for some other purpose, however good and well-considered. The quaintly named Office of Scientific Integrity (OSI) is similarly inviolate. It represents an earlier compact between the NIH and Congress that left with NIH responsibility for the self-regulation of

its own cases of misconduct in science; the promise that OSI would at the same time enjoy a degree of autonomy within NIH was implicit in that deal. Healy has been required publicly to acknowledge that that is the case.

Needs

Yet her instincts are correct. The constitution of OSI as it stands is thoroughly unsatisfactory. Given the pervasiveness of NIH grants in US research, it has become *de facto* the ultimate decision-making body on all allegations of scientific misconduct in biomedical research. But there are two distinct and separate needs, only awkwardly met by OSI, as illustrated by the two best-known cases now on its books. First, NIH need a system for investigating allegations of misconduct in their own intramural research (typified by the case of Dr Robert Gallo and the identification of the AIDS virus). The NIH collectively have the resources to investigate questions such as these quickly and, where possible, decisively. The natural course would be for the director of NIH to appoint an external committee of investigation and to promise whatever resources are required. It is difficult to believe that, had such a course been followed in the Gallo case, the investigation would have dragged on for the three years OSI has so far spent on it.

Extramural cases are different. OSI has status in them only because NIH money, in the form of grants or

BRITISH SCIENCE

We shall be publishing on 12 September *Nature's* Manifesto for British Science, which will be offered without charge to any one of the political parties contesting the next general election, and in the belief that it is an improvement on their own version.

Following publication, on Friday 13 September, there will be a public meeting at the Royal Institution, Albemarle Street, London W1, at which Sir Mark Richmond, chairman of the Science and Engineering Research Council, will be the principal speaker. The chairman will be John Maddox, editor of *Nature*. The meeting will start at 9.45 a.m. and end at 12.45 p.m. The political parties will be invited to send representatives, but there will be ample time for discussion.

Those wishing to attend should apply for tickets to Mary Sheehan, *Nature*, 4 Little Essex Street, London WC2R 3LF. □



contracts, is involved, but first responsibility for investigation of alleged misconduct rests with the recipient institution. Were it not that institutions have shown themselves in the past few years often (but not always) incapable of treating allegations of misconduct as seriously as they deserve to be, OSI's role in cases such as that surrounding Dr Thereza Imanishi-Kari's research at MIT would have been much less direct. If the research community wishes to get OSI off its back, it should ensure that mechanisms by which institutions investigate allegations of misconduct are more incisive. Sadly, professional societies in the United States seem embarrassed by the whole business, and nothing has been heard from the US National Academy of Sciences since the beginning of the Imanishi-Kari business.

Instincts

Healy's instincts would also be correct if she were offended by the way in which OSI functions. There is ample reason to believe that people against whom misconduct is alleged may be unjustly pilloried. OSI's investigations are conducted in camera and those accused of misdemeanours are not legally represented. OSI's justification of this procedure is that the matters to be decided are largely scientific, not legal. (Even so, OSI is more inclined to rely on documentary rather than laboratory evidence, when it would often be helpful — and quicker — to commission replications of disputed experiments.) But what if the outcome is the ruin of a person's career in research? Can it be fair that what is called in the United States "due process" should be so waywardly denied? There might be some excuse if OSI could offer speed instead, but that is hardly the case. Sadly, OSI's proposals for modifying its procedures do not promise every change that is required. At some stage, there must be a more radical reform.

Healy, despite last week's setback, is best placed to bring that about, but only with the support of the Congress and of Dingell in particular. But is not Dingell the arch-villain, the congressman most determined to bring the research community to heel? In the synthetic demonology of the past few years, that may be how he has been made to seem, but there is no reason to believe that Dingell is moved by an animus against research. There is also no reason to suppose that he would not listen to a reasoned case from Healy for the reform of OSI and its repositioning within the NIH system.

What form should such a proposal take? Analogies with the courts are unavoidable, and OSI should become a kind of appeal court to which all parties to allegations about NIH-funded work would have access — whistleblowers as well as those accused of misconduct. OSI itself should consist primarily of a panel of scientists respected for their integrity, but also for a healthy realism about the ways of the real world, who are willing to devote a substantial part of their time to these matters.

Access to this remodelled OSI should be open to all,

but should initially be confidential. NIH should offer whatever scientific resources are needed for investigations even of extramural cases to be carried out. As things are, OSI has relied too much on documentary evidence; in the Gallo case, for example, it would have been preferable that OSI, rather than Gallo himself, should have expeditiously investigated the nucleotide sequence of samples of virus recovered from early material recovered from his deep freeze for evidence of the provenance of authentic HIV. But having reached a conclusion from which some of those concerned dissent, OSI should be prepared to defend its position publicly, in what would essentially be a quasi-judicial process in which all those concerned and their lawyers would be entitled, if they chose, to have a say. □

Hedging the issue

The British government seems about to end decades of wrangling over the fate of Britain's hedgerows.

SINCE the 1950s, successive British governments have seemed unable to decide whether farms should be given grants to maintain and develop the hedges on their property, or grants to rip them up. Until 1985, the answer was "both". In that year, the government bowed to pressure from groups such as the Council for the Protection of Rural England and ended the funding of hedgerow demolition. The argument that such destruction was necessary for the efficient exploitation of agricultural resources no longer seemed tenable in the face of Europe-wide food surpluses.

The government's white paper (policy document) *This Common Inheritance*, published earlier this year, aimed to continue the trend towards preservation. Under the proposed scheme, farmers intending to remove a hedgerow would be obliged to notify their local authority, which would then decide whether the threatened hedgerow was of sufficient wildlife, historical or aesthetic value to warrant preservation. If so, then the farmer would be issued with a grant to pay for the upkeep of his small corner of the national heritage.

It did not take long for conservation groups to point out that all a shrewd farmer would have to do would be to declare his intention to tear down all his hedges to be certain of at least one local authority grant for the upkeep of a hedge whose maintenance had previously been a negligible part of his annual expenditure.

In the light of this piece of common sense farmers will no longer automatically qualify for a maintenance grant if the local authority decides that a hedge they want to pull down is worth saving. As Mr Tony Baldry, the junior minister at the Department of the Environment, puts it: "In some cases the most environmentally beneficial course of action may be to leave a hedgerow in an unmanaged state for the immediate future." □

Stanford counts cost of overhead scandal

- President Donald Kennedy resigns
- Cuts loom in programmes and staff

Washington

THE biggest news out of embattled Stanford University last week was that president Donald Kennedy announced he will retire at the end of the coming academic year. Kennedy said he had become too closely associated with the problem of inflated overhead costs at Stanford to be an effective part of the solution.

But even as Stanford was coming to grips with Kennedy's resignation, another consequence of the indirect costs fiasco was beginning to take shape — one that will affect the university's future much more dramatically than the loss of any one individual. Faced with a major decline in revenue, brought on in large part by cuts in its indirect costs rate, the Stanford administration has said it will get rid of those programmes that are least valuable to the university.

Understandably, faculty and staff morale has slumped, as university employees wait for the axe to fall. The administration, however, is trying to make the best of a bad situation by taking this opportunity to restructure the university — perhaps in a major way. "They're asking what they'd do if they started from scratch," says dean of research Robert Byers. It is a time for soul-searching at this major research university, and five years from now Stanford could be a very different institution.

The funding crisis has been coming for several years, and many of its roots lie in an ambitious building programme Stanford undertook in the 1980s. As expenses increased, the indirect costs rate charged by the university on federal grants also rose until it became one of the highest in the country at more than 70 per cent. Many inside and outside the university have claimed that Stanford tried consciously to charge as much as possible to indirect costs — a practice that led eventually to such well-publicized excesses as the inclusion of flowers for a wedding reception and depreciation on a yacht on the overhead bill.

These abuses (and less colourful ones from other universities) were aired in a series of hearings of a House of Representatives subcommittee chaired by John Dingell (Democrat, Michigan). The upshot was that Stanford returned some \$1.3 million that it had overcharged the federal government for indirect costs, and the federal government trimmed Stanford's indirect cost rate back to 55.5 per cent, pending new negotiations.

Citing the stain on Stanford's reputation that the indirect costs furore produced, Kennedy resigned. "It is very difficult, I have concluded, for a person identified with a problem to be the spokesman for its solution," he said.

Many of the faculty applauded his decision, and the belief was widespread that although Kennedy had done a fine job in the 1980s, his effectiveness is now mostly gone. "Kennedy served this insti-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Kennedy — too close to the problem.

tution well in many ways," said one researcher, "but he got himself identified with a number of problems and his presence became divisive on campus. I think he should have [resigned] earlier."

Kennedy is leaving Stanford to face a large deficit, created mostly by the shrinkage in indirect funds. To balance the books, the university will have to cut about 10 per cent from its operating budget of approximately \$400 million a year. That comes on top of \$22 million in permanent cuts it took from the administration and support portion of its operating budget over the past year.

The additional \$40 million the university must carve out of its operating budget will make it necessary to get rid of entire programmes. "This is not going to be an exercise where we're going to do across-the-board cuts," says Robert Freelen, vice president for public affairs. "To preserve Stanford's quality, we will have to make judgments about size and quality" of individual programmes.

At this point, no one knows where the cuts will come, although a tentative schedule for the cuts has been set — \$20 million in the 1992-93 academic year, followed by \$14 million and \$6 million in the next two years. The administration has asked each department to decide

what should stay and what should go if the department were to be trimmed by varying amounts up to 20 per cent.

Although Byers says this effort "is not focused on the cuts, it's focused on setting priorities", the faculty is still nervous. "The wolf is at the door, but he hasn't actually had a meal," is the way one researcher describes it.

The cuts could be debilitating, says Bob Simoni of the biological sciences department. His department has been asked to plan for a cutback of 6 per cent on operating budget, he says, but half of those expenses are for salaries, which cannot be cut, so the department must figure out how to survive with a loss of 12 per cent from its non-salary budget.

Among some faculty members, there is a feeling of resentment that the administration, which got the university into its present mess by bumping up Stanford's indirect cost rate during the 1980s, is now asking for cuts in research programmes. These researchers want Stanford's administration — which some portray as a bloated bureaucracy — to take the brunt of any cuts. But, as Theodore Geballe of the applied physics department points out, administrative cuts can make up only a part of the savings required to balance Stanford's budget — cuts in faculty positions are inevitable. Byers adds that, contrary to the grumbling of some faculty members, the administrative side is indeed scheduled to take a greater percentage of the cuts than the faculty.

One result of the financial crunch has already become manifest: Stanford's ambitious growth plans have ground to a halt. A planned \$250 million programme to redevelop the university's Near West Campus, a 41-acre science and engineering region, has been scaled down considerably and postponed. Administration officials say that the west campus plan is not dead, but no one is predicting when it will get restarted.

To cope with its financial crisis as best as possible, Stanford is making the cuts a part of a restructuring of the entire university. "Nothing is sacred," Byers says. Not only will some programmes be abandoned entirely, but the organization of others could be shuffled to create new departments that cross traditional boundaries. The university has formed several committees from both the faculty and the board of trustees to make recommendations about the future shape of the school, and their initial recommendations will be offered over the coming month.

"We're preparing the university for the next century," Byers says. "It's an exciting time." Exciting perhaps, but some faculty members would gladly settle for a little less excitement and a little more stability.

Robert Pool & Peter Aldhous

The perils of privatization

London

It seemed too good to be true. In the spring of 1988, then Prime Minister Margaret Thatcher announced plans to extend her administration's strategy of privatizing state-run enterprise by splitting off the national electric utility into four competing companies. This, it was argued, would lower prices while retaining the utility's large environmental and nuclear research effort, thanks to a disproportionate division of assets that would leave one of the companies — National Power — wealthy enough to invest in such long-term projects.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

No place for research: profit rules the power plants.

Three years later, after a last-minute compromise with parliament that stripped National Power of much of its advantage, it is clear that Thatcher's privatization plan was, indeed, too good to be true. The utilities are now in fierce rivalry and National Power has announced its intention to cut its overheads by a half in an effort to reduce costs and remain competitive. Although officials will not reveal the details, company scientists say that National Power is preparing to dismantle most of the £26 million in non-nuclear research that it inherited from the Central Electric Generating Board, as the electric utilities were collectively known when they were under the government's wing.

Scientists at National Power's Leatherhead laboratory say that all but about 50 to 70 members of the 500-person research staff will be laid off later this year, and that the remainder will be transferred to the company's Swindon offices. This will effectively mean the end of one of the world's largest acid rain teams, as well as those teams that researched other environmental effects of coal-fired electricity generating plants, from sulphur dioxide to ground-level ozone. Other groups that expect to be eliminated include a superconductivity team and researchers who study the effects of power plants on the nearby bodies of water that serve as source and dump for coolant.

In a letter in June to Labour Party

energy spokesman Rhodri Morgan, John Baker, the head of National Power, explained that the company did not intend to maintain "research programmes which were addressed to the solution of national problems or dealing with national or international issues where these extend far beyond the boundaries of our own company interests." In short, National Power wants to cut any research that does not directly help its cause, even if it is of national importance. Research in the public interest is for the government to do, say company officials, not them.

So what went wrong? Ultimately, the government's desire for real competition in the electric industry sealed the fate of the research. Unlike previously privatized industries such as British Gas and the water utilities, which were all essentially given monopolies when they were sold, the electric utilities were purposely set up to fight for customers. The competition is even closer than originally intended. Although Thatcher wanted National

Power to take over the nuclear power industry and run it privately (which would have given the company some 70 per cent of the UK energy market), parliament forced her to choose between privatization and expansion of nuclear power. She chose privatization, and returned nuclear power to the government, leaving National Power with a slim 20 per cent lead over the second-largest privatized utility, Powergen.

"We had no idea that National Power would be shrunk so quickly and shorn of the mindset that it was operating for the public good," says Morgan. Under stockholder pressure to trim costs, National Power has acquired an "overwhelming desire to demonstrate that they can be more private sector than the private sector," he says. "In something as environmentally sensitive as electricity generation, this is a real mistake."

Anticipating that the new battle lines would put a premium on rock-bottom pricing — to the probable detriment of research — parliament inserted a clause in the 1989 privatization act that gave the secretary of state the power to direct the utilities to do research "in the public interest", says Morgan, who helped to write the legislation. "The 'acid rain test' is whether the secretary of state actually uses the provision. We'll ask him to, but he'll probably fight like hell."

So far there are no signs that the government intends to intervene. Mor-

gan says that is because National Power has not officially announced the specifics of its cuts yet. But others expect little action even after the cuts are enumerated.

"It's almost impossible to define a 'national interest', especially when it comes to science," says Robert Preddie, a former engineering director at the Central Electric Generating Board and a member of the independent Institute of Energy. "This government's interest in research is minimal — the issue wouldn't even get to square one."

Just how much the government cares about what was once its 'national interest' energy and environment research will be put to the test this autumn, when the House of Commons plans to hold an Energy Select Committee inquiry on the consequence of the electricity privatization. Research — and the lack thereof — is expected to be a major topic.

Christopher Anderson

ENVIRONMENT

Europe's green thumb firmly on nose

London

Member states were caught violating European Communities (EC) environmental law almost 170 times last year, 66 per cent more than the year before and more evidence that the Communities' environmental legislation is widely flouted by its members (see *Nature* 352, 183; 18 July 1991). Of the total of 167 violations, 131 were for a failure of the member states to pass national legislation in conformance with EC dictate, almost three times the previous year's figure, according to a report released last week. But EC officials explain the soaring rate as a sign not of decreasing compliance but of increasing vigilance of regulators in tracking down and prosecuting offenders.

Of the identified cases of violation, only 39 resulted in formal legal proceedings; the rest were just letters of warning. But that is still some 53 per cent more than in the previous year and the highest level of the decade. Of last year's cases, only 16 (down five from 1989) reached the European Court, a sign that member states are beginning to back down when challenged.

To speed the trend, EC officials are beginning to use reports such as this one to shame countries into compliance. In recent years, the EC has taken to publishing community-wide data books, partly as a way to embarrass the countries that refuse to submit their own measurements. Last year, an EC report on bathing water data had nothing but blank pages where Germany and Greece should have been, which — after some media outrage — prompted those countries to institute reforms.

Christopher Anderson

Healy and Dingell lock horns

Washington

BERNADINE Healy last week weathered her first serious storm since her appointment as director of the US National Institutes of Health (NIH), going on the offensive in a series of angry exchanges over her management of the NIH misconduct office with members of the House of Representative investigations and oversight subcommittee chaired by John Dingell (Democrat, Michigan).

But while Healy may have emerged relatively unscathed from the hearing, it seems that the troubled NIH Office of Scientific Integrity (OSI) may be in for further upheaval. Dissatisfied with Healy's moves to reform the office since her arrival at NIH, Dingell is expected to recommend that OSI be moved out of NIH, and into the Inspector General's office at the Department of Health and Human Services.

Dingell's subcommittee wanted to explore Healy's reorganization of the OSI, which has already led to the resignation of former OSI deputy director Suzanne Hadley from the office's two most prominent misconduct investigations (see *Nature* 352, 268; 25 July 1991). Some have suggested that Healy's manoeuvres might be linked to the fact that a Hadley-authored draft report on alleged misconduct at Healy's former institute, the Cleveland Clinic Foundation, would reflect badly on Healy (see *Nature* 352, 361; 1 August 1991).

In May 1990, Healy presided over a preliminary inquiry at the clinic into allegations that a Cleveland biochemist had made false claims in NIH grant applications. Her panel concluded that the researcher had, indeed, described experiments that had not been performed, but characterized this as an honest error rather than wilful misconduct: the researcher had neglected to remove 'anticipatory writing' about experiments he had expected to perform from the final grant application.

During last week's hearing, Healy admitted freely that the preliminary Cleveland inquiry was inadequate — among other irregularities, a co-investigator on the disputed grant was included on the inquiry panel. But time and time again, Healy insisted to the congressmen that they should not focus exclusively on her actions in that initial inquiry. She had personally instigated a second inquiry, she pointed out, and when the second panel expressed grave doubts about the earlier decision, she ensured that no money under the disputed grant was ever spent.

But Dingell and his colleagues were determined to focus on the first inquiry, and tempers became steadily more

frayed. "I can't handle this witness," said Representative Norman Lent (Republican, New York), at one point.

Healy also defended her handling of OSI. She said her review of OSI was influenced by serious breaches of confidentiality in the office, notably the widespread leaking of a draft OSI report on the Imanishi-Kari case and the accidental mailing of a confidential tape from the Gallo investigation to the lawyer of Gallo's former colleague Mikulas Popovic.

She said it was "preposterous" to suggest that the continuing OSI investigation into the Cleveland Clinic case had biased her handling of the office and of Hadley. She stood by her decision to ask OSI director Jules Hallum to retrieve Hadley's files on the well-publicized investigations into the conduct of National Cancer Institute AIDS researcher Robert Gallo and Tufts University immunologist Thereza Imanishi-Kari. Because Hadley had left her position as a full-time investigator at OSI in March but had continued to work on the Gallo and Imanishi-Kari files from another office in NIH, Healy said the result was that a "quasi-independent, satellite operation" had been set up which communicated poorly with the rest of OSI.

Despite her insistence that she has done nothing wrong in her handling of OSI, Healy has now retreated from active involvement with the office, passing responsibility to Carl Kupfer, NIH's acting director of intramural research, until OSI has released its final report on the Cleveland Clinic affair.

In a telephone interview, Healy questioned the proposal that OSI should be moved out of NIH. Dingell's concerns, she suggested, are based on the misinterpretation that OSI is the final adjudicator in cases of alleged misconduct. Despite the impression given by the leaking of the draft Imanishi-Kari report, the final judgement rests with assistant health secretary James Mason, Healy noted.

It seems unlikely that Healy has heard the last of the current OSI controversy. A statement from Bruce Singal, lawyer for Imanishi-Kari, saying that Healy "has had the courage to insist on conducting this investigation so that Dr Imanishi-Kari's constitutional rights are safeguarded" may serve to embroil Healy further in a controversy in which the views of the research community are becoming ever more polarized.

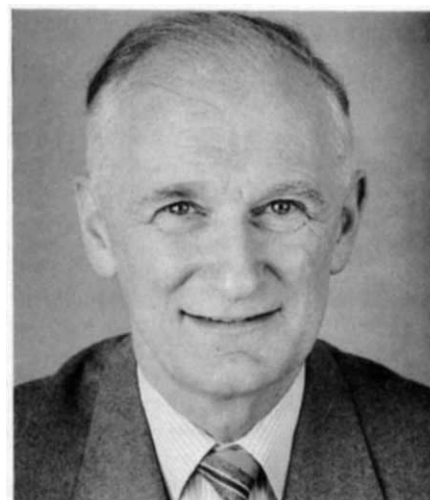
Healy seems resigned to a difficult path ahead. Hers is a job in which there are a lot of obstacles, she said. "If I'm to be effective, I guess I'm going to have to learn to work in that environment."

Peter Aldhous

New head to unify schools, shed science

London

In a first step towards the planned unification of the UK university, polytechnic and college systems, the agencies that run the three sectors will now have the same chairman. Sir Ron Dearing, currently the head of the Polytechnics and Colleges Funding Council, was appointed last month to head the Universities Funding Council as well. High on his agenda when he takes his new post on 1 November, will be the unification of the two councils and transfer of science funding to the research councils, which are both part of an overall reorganization of UK higher education management.



Dearing will get all of the schools and none of the research headache.

As laid out in a letter to Dearing from Education Secretary Kenneth Clarke last month, the two funding councils are to merge in April 1993. Rather than distributing council responsibilities by type of higher education institution, as is now done, unified funding councils in Scotland, Wales, Northern Ireland and England will be assigned all higher education in their geographic areas.

Unlike his predecessors, Dearing will not directly control research funding. Beginning on 1 August, 1992, the education ministry will no longer pay for the overhead portion of research grants funded by the research councils, in part to avoid the sort of accounting nightmares that have recently plagued US universities and funding agencies (see story, page 459). Rather than negotiate indirect cost rates for each institution, as US agencies do, the research councils will add a flat 40 per cent to all research grants to cover indirect costs, Clarke announced last week. The government will shift some £150 million in its 1992-93 budget from the education ministry to the research councils to cover the transferred responsibilities, he said. Christopher Anderson

Gyro failures worry NASA

Washington

THE failure of two of the six gyroscopes that keep the Hubble Space Telescope pointing at targets under observation is causing a major headache for the US National Aeronautics and Space Administration (NASA). Further failures could force the agency to launch an emergency space shuttle mission to replace Hubble's gyros, before the first planned shuttle visit in late 1993.

The Hubble telescope was designed to use four gyros to provide accurate positioning for its scientific observations, with two backups. In addition, Hubble has three coarser backup gyros, which cannot position the telescope with sufficient accuracy to yield useful science, but should at least keep the satellite from spiralling out of control if the science gyros give out.

NASA officials were surprised but not too concerned when one of the science gyros failed last December. But a second gyro failure on 30 June has forced project managers at the Goddard Space Flight Center to produce emergency plans to rescue the telescope in case of further problems. Since then, a third gyro suffered a momentary glitch on 26 July, and although it is still working, its characteristics have changed subtly, says John Campbell, deputy associate administrator for Hubble operations at Goddard.

After the 30 June failure, NASA officials decided to run Hubble on three science gyros, rather than turn on another backup. "To our surprise, we were able to operate very well," says Campbell. One early theory was that the high solar activity since Hubble's launch could be to blame, which argued against turning on the last backup science gyro until solar activity died down. But Campbell says that NASA's experts have now ruled out solar radiation as a factor in the failures, and the last science gyro has been turned on.

Because NASA officials cannot control Hubble with two science gyros, the satellite now stands only two gyro failures away from becoming scientifically useless. After that, the failure of just one of the three coarser backup gyros could cause the telescope to be lost altogether. NASA software engineers are now working frantically on a computer program that would allow Hubble to be kept under control in this eventuality, by using the telescope's solar and gravity sensors to keep it pointing perpendicular to the Sun.

The possibility of an emergency shuttle mission to fix the gyro problem is already under active discussion. A shuttle flight is planned for 1993, to replace the telescope's solar arrays and the Wide Field/

Planetary Camera, and to install corrective optics in front of three other instruments at the base of the telescope. But little of this equipment will be ready until just before that mission, so if an earlier flight to fix the gyros becomes essential, a second visit would still be necessary.

No decision on fixing the gyro problem can be made until the reasons for the failures are identified, and, so far, no common factor to explain the two failures has been found. The December failure

US NATIONAL LABORATORIES

California's record attacked

Washington

THE University of California's management of three national laboratories came under stinging attack in Congress last week, just a week after the Department of Energy (DOE) had decided to retain the California system as contractor to run the facilities (see *Nature* 352, 365; 1 August 1991). Among the litany of problems trotted out before a House subcommittee were the loss of \$18.6 million worth of equipment and inadequate accounting for hundreds of millions of dollars in indirect research costs.

The three laboratories, Lawrence Berkeley and Lawrence Livermore in California, and Los Alamos in New Mexico, have all been managed by the University of California for more than 40 years. But concern about the adequacy of the university's management of the laboratories had been raised in a series of reports from the congressional General Accounting Office (GAO) and by the Inspector General at the DOE, fuelling speculation that the department would seek a new contractor — at least for Lawrence Livermore and Los Alamos, home to sensitive nuclear weapons research.

This speculation was quashed by last month's decision of Secretary for Energy James Watkins to begin negotiations to extend the university's contracts. But Howard Wolpe (Democrat, Michigan), chairman of the investigations subcommittee, said that the DOE must seek changes in the university's contracts to ensure better management in future. "If the Department is unwilling to publicly commit to securing . . . these critical changes in its contracts, then it has undermined its own position going into negotiations." DOE officials say they cannot discuss details of the contracts while they are being negotiated.

Victor Rezendes, from the GAO, told the subcommittee that the university has given higher priority to defending its own patent rights than to protecting the

has been traced to the electronics designed to return the gyro back to the centre of its range of movement, after being disturbed. But the second gyro failure seems to be linked to the electronics that apply power to the gyro wheel — a seemingly unrelated problem.

For Campbell, the difficulty in pinning down the problem with Hubble's gyros is a little disturbing: "If you can't find a generic flaw in the electronics, you have to assume that it was badly built. It's like a car — if you have a lemon, it's difficult to know what to do."

Peter Aldhous

federal government's interest in defence-related technologies developed by Livermore researchers. Because in-house patent attorneys at Livermore were busy on university business, the DOE's San Francisco field office had to hire outside lawyers during 1990 to patent a laser technology called U-AVLIS (uranium-atomic vapour laser isotope separation), used in uranium enrichment.

In addition, around 100 researchers at Lawrence Berkeley were employed on research paid for by National Institutes of Health (NIH) grants, Rezendes said. Wolpe said that a letter from NIH director Bernadine Healy had informed him that much of this work did not necessarily require the facilities available at Lawrence Berkeley. Under the rules of the DOE's 'work for others' programme, national laboratories should not accept outside grants when the work could be equally well carried out in private sector laboratories.

Asked about the accounting for indirect research costs, which amounted to some \$350 million at Livermore in 1987 (more than 35 per cent of the laboratory's total funding), DOE Inspector General John Layton said "we don't know how the money's being spent".

University officials declined the invitation to testify before the subcommittee. James Kane, who has headed the university's recent discussions with the DOE, says that the university had been willing to address the specific issue of the non-standard clauses in its management contracts. But the scope of last week's hearings was so broad that it would have been difficult to prepare an adequate testimony, he says. Kane maintains that the University of California's contracts are broadly similar to those held by the DOE's other non-profit-making contractors — the university associations that run Fermilab and the Brookhaven National Laboratory, and the University of Chicago, which runs the Argonne National Laboratory.

Peter Aldhous

A leap into controversy

Washington

A CALIFORNIA-BASED television scriptwriter has found himself at the centre of a bitter row between biomedical research advocates and animal rights activists, who have learned of his plan to feature a chimpanzee involved in a head injury research project as the central character in an episode of a popular US television series.

Paul Brown writes for the NBC-networked show *Quantum Leap*, in which the hero is a scientist of the future who transports himself into the bodies of characters from other times. Previously, the show has used this format to explore controversial issues such as racial discrimination. But the storm whipped up by a report in the national television listings magazine *TV Guide* about Brown's plan to have the character appear in the body of a research chimpanzee has surpassed anything the show's executives have experienced before.

That report quoted Deborah Pratt, co-executive producer of *Quantum Leap*, as saying "the animal rights people are in heaven," and has prompted a furious letter-writing campaign by the California Biomedical Research Association and a host of individual researchers. "It appears that in the development of this episode, the lies and misrepresentations of the animal rights movement have been accepted," says one letter to *Quantum Leap*'s producers, from David Ramsay, the University of California, San Francisco. Sandra Bressler of the California Biomedical Research Association argues that script should perhaps include a head injury patient helped by results from the

research project.

Animal rights groups respond that Bressler and her allies, by putting pressure on the show's producers, are attempting to suppress free speech. Shirley McGreal from the International Primate Protection League says she fears that the pressure being applied by the biomedical research lobby will result in the programme being cancelled.

The script, set in 1972, involves a chimpanzee used for language research, which is then transferred to a project investigating head injuries. The chimpanzee receives a number of blows to the head and is killed, after which the project leader discovers that the chimpanzee is not an appropriate model to study head injuries in humans.

Brown's idea is based loosely on a head injury research project described in the *Journal of Neurosurgery* (39, 152; 1973). Ayub Ommaya, the principal investigator on that project, says the chimpanzee work was a one-time experiment to check that work with smaller laboratory animal models could be scaled up to apply to human head injuries. The project yielded data that were used to develop restraints to minimize head injuries in road accidents, he says.

Brown's characterization of a similar research project as being unnecessary annoys Bressler. "I appreciate artistic license, but . . . when you're dealing with an issue of such significant controversy, you have some obligation not to misrepresent it to one side or the other." But Brown is unperturbed by criticism of his script. "Good television will always stir up controversy." **Peter Aldhous**

BIOMEDICAL RESEARCH

New institute snares prize team

Washington

A NEWLY established biomedical research institute in Long Island, New York, has made a good start by luring a leading research team from Rockefeller University in New York City. Anthony Cerami, whose work has centred on cytokines and on diabetes, and 13 of his colleagues will begin work later this summer at the Picower Institute at Long Island's North Shore University Hospital.

The institute has been given an initial endowment of \$10 million by the financier Jeffrey Picower and his wife Barbara, but Cerami says he expects this to grow tenfold over the next five years, allowing the institute to house some 100 researchers and to rival in size such established centres as the Whitehead Institute in Cambridge, Massachusetts.

Approximately 80 per cent of the new

institute's work, Cerami says, will be research into diseases associated with ageing.

Cerami's departure from Rockefeller has attracted considerable publicity because he was prominent among a faction of the Rockefeller faculty that two years ago argued against the appointment David Baltimore as Rockefeller president. Cerami feared that Baltimore's association with the controversy surrounding the alleged fabrication of data by Thereza Imanishi-Kari in a disputed 1986 *Cell* paper would undermine his ability to serve as Rockefeller's president. But Cerami says his decision to leave the university was influenced by the unique opportunity given by the Picowers' offer, rather than by Baltimore's presence at Rockefeller. "I would have left no matter who was president." **Peter Aldhous**

Fullerenes heat up

Washington

A research team led by Ray Baughman and Zafar Iqbal, from Allied Signal's laboratories in New Jersey, last week announced that they have produced a superconducting doped fullerene with a transition temperature around 42 K. The finding, revealed at a meeting in Philadelphia, exceeds the previous highest temperature claimed for fullerene superconductivity by some 10 K. The compound is doped with rubidium and thallium, and Baughman says that adding thallium to both rubidium- and potassium-doped fullerene seems to raise the transition temperature by several degrees. Rick Smalley, a leading fullerene researcher from Rice University in Houston says, "The consensus of people at the meeting was that it is a very impressive result, and probably real". **P.A.**

Switchboard in orbit

AFTER the usual series of false starts, the US space shuttle Atlantis carried a Tracking Data Relay Satellite (TDRS) into orbit last Friday, 2 August. The TDRS network of geostationary satellites is vital for communication between US space science satellites and National Aeronautics and Space Administration (NASA) ground stations — each satellite acts as an orbiting communications 'switchboard', relaying data from missions such as the Hubble Space Telescope and the Gamma Ray Observatory back to Earth.

Until last week, NASA had three of these satellites in orbit. But two of the three are not working properly, and are being run together to mimic a fully functional TDRS. With one of these satellites nearing the end of its expected life, last week's TDRS launch was a high priority, to ensure good data communications for NASA's planned science missions. **P.A.**

Space exchange

THE improved relations between the United States and the Soviet Union will soon be reflected in the first US-Soviet cooperation in manned spaceflight since the symbolic docking between Apollo and Soyuz capsules in 1975. During their meeting in Moscow last week, Presidents George Bush and Mikhail Gorbachev agreed to send a US astronaut to the Soviet Mir space station for several months, and for a Soviet cosmonaut to fly on a future space shuttle mission.

The missions are expected to concentrate on life sciences and medical research — now the chief scientific justification for the \$30,000-million-plus space station Freedom project. A priority will be standardizing the different measurements used to monitor the health of astronauts under weightless conditions. **P.A.**

Conditional approval touted

Washington

AIDS patients in the United States, often with the approval of their doctors, are going to the black market to get drugs that have not been approved by the Food and Drug Administration (FDA). The solution, many argue, is for the FDA to offer a conditional approval, in which drugs could be marketed on the basis of preliminary safety and efficacy data on condition that they would be reassessed when the clinical trials usually required for FDA drug approval are complete.

Marcus Conant of the University of California in San Francisco points to experience with the anti-retroviral drug DDC. Combined therapy with DDC and zidovudine (AZT) retards the loss of CD4 cells — the key to the long-term survival of people infected with human immunodeficiency virus (HIV) — more than either drug alone, according to results of a clinical trial conducted by Margaret Fischl of the University of Miami. But DDC is not approved by the FDA, so a patient's only legal means of obtaining combined DDC/ AZT therapy is to enrol in the appropriate clinical trial. Space is limited in these trials, however, and not all patients meet the entrance requirements, so many AIDS patients go to the black market for DDC. "If I were in their shoes I would be doing the same thing," Conant says.

Besides being illegal, buying DDC on the black market has another drawback: patients are not assured a continuous supply of the drug. The telephone of one black-market DDC outlet in San Francisco has a recorded message that the drug is unavailable until the middle of August; then it will be sold on a first-come, first-served basis.

NUCLEAR LABORATORIES

Radioactive frogs

Washington

FROGS' legs are definitely off the menu at the Oak Ridge National Laboratory, Tennessee. After a bumper breeding season, scores of leopard frogs, *Rana pipiens*, have started to migrate from two ponds at the site. The trouble is that, in the 1940s and 1950s, these ponds were used for the storage of intermediate-level liquid radioactive waste. After one frog was run over on the site and found to be contaminated, the laboratory's health physics staff closed a stretch of road near the ponds and are sampling other frogs on the site for contamination. Those that are clean are dropped into an uncontaminated creek; those that are not must go back to the contaminated ponds. Some staff at Oak Ridge became alarmed when one frog was found in the basement of a laboratory

Conant calls for the FDA to speed the approval process by instigating conditional approval, and the agency says it is currently evaluating various means to do just that. It is expected to release a preliminary report on the subject sometime in the next few weeks.

Last month, the FDA's Antiviral Drug Products Advisory Committee recommended that the agency approve for sale the anti-retroviral drug DDI on the basis of limited data from phase 1 (safety) and preliminary phase 2 (efficacy and side effects) clinical trials (see *Nature* 352, 269; 25 July 1991). Although the committee did not formally recommend conditional approval, it did ask the FDA to reassess the drug when the phase 2 trials are complete — a sort of *de facto* conditional approval.

■ Meanwhile, Conant reports a surprising clinical observation. Twelve AIDS patients have survived for between three months and two years with negligible numbers of the supposedly indispensable CD4 lymphocytes, he says.

Most people with AIDS die of opportunistic infections, such as *Pneumocystis carinii* pneumonia, within one year of their CD4 counts falling below 50 cells per cubic millimetre of blood. But the 12 patients' CD4 counts have hovered around zero for the past year, during which time they have never exceeded 10. A healthy person has a CD4 count of 800–1200. "A few years ago we would have said that life was incompatible with zero CD4 cells," Conant says.

The key to the survival of patients lacking CD4 cells is early diagnosis and aggressive preventive therapy against opportunistic infections, Conant says.

Rachel Nowak



office building, but Kornegay says there is little to worry about. The frogs are not contaminated externally, and could pose a threat to health only if eaten. The species is not edible.

P.A.

Science goes video

Washington

GARY Welz has an ambitious plan. By January 1993, he hopes to launch the world's first television network aimed specifically at working scientists. And last month he received a big boost towards that goal. After a year trudging the corridors of various US scientific societies and grant-making bodies with his proposal, the mathematician-turned-video producer received some financial backing to begin work on his project, in the shape of a \$30,000 grant from the Alfred P. Sloan Foundation.

The network would offer a mixture of news bulletins, televised scientific conferences and live interviews with researchers, Welz says, and it would be supported by income generated through advertising. It would be a service not dissimilar to that provided by the two leading multidisciplinary scientific journals, *Nature* and *Science*, except that it would be in video rather than printed form.

That parallel is not lost on Welz. When he explains his idea of raising additional funds by having some companies pay for feature-length programmes about their research activities, he notes that *Nature* has set the precedent, with the publication earlier this year of a supplement featuring the French oil and petrochemicals company Elf-Aquitaine.

Initially, Welz hopes to broadcast for a few hours a week, possibly via cable in centres of intense research activity, such as Cambridge, Massachusetts, and Palo Alto, California. But \$30,000 is only a small fraction of the funds needed to put a television network on the air, and Welz has yet to provide convincing proof that there really is a demand for his product. He believes there is a US audience of more than 500,000 — the 350,000 people who have earned science and engineering PhDs in the United States since 1960, some 120,000 current graduate students, as well as crossover interest from US physicians. But neither Welz nor anyone else can say just how many of these researchers — most of whom have difficulty keeping abreast of the literature in their fields — would be interested in adding a science television channel to their agendas.

Tom Wolzien, vice-president for cable projects at NBC television, says the initial problem for narrowly focused projects such as Welz's proposal is obtaining distribution. Before cable television operators will agree to distribute a new channel, he says, they need convincing that it will bring in advertising or subscription revenue. Welz, who is based at the City University of New York, is now busy writing further grant applications to secure funds for the all-important market research.

Peter Aldhous

Cooperation in Hong Kong

Hong Kong

UNLIKE most countries in the West, Hong Kong does not have a history of cooperation between industrial companies and academic institutions, something that has proved to be a drag on the country's competitiveness in high-technology business. But that is now about to change.

Six publicly funded tertiary institutions in the territory have just set up a non-profit company, Varsity Resources Corporation Ltd, whose aim is to channel academic expertise into the private industrial world. Not only will it allow industrial businesses to improve their research and development skills, but it will provide Hong Kong researchers with a practical outlet for their expertise.

Other countries in the Far East have found various ways to promote such industrial/academic cooperation, but it has usually been done with the services of a political body as a midwife. Taiwan's government, for example, has poured millions of dollars into science parks close to universities and institutes of technology. The government of Singapore has created a 'technology corridor' in which small companies are invited to set up shop close to the national University of Singapore.

Hong Kong is different. Over the years, its *laissez-faire* government has opposed anything that smacks of an industrial policy. According to the Hong Kong government's philosophy, the market is the only effective selector of technological winners and losers. Funding for research and development amounts to no more than 0.04 per cent of Hong Kong's gross domestic product (GDP), a low figure even by Asian standards. By comparison, Western nations such as the United States and Germany spend 2.5-3 per cent of their GDP on research.

Over the past two years, a few Hong Kong companies have become increasingly aware that the colony's industrial sector — which consists predominantly of companies with fewer than 50 employees — lacks some crucial research and development skills. As a result, the territory's technological capacity is falling behind those of Korea, Singapore and Taiwan, the other three 'dragons' that have created strong technology-based economies.

About 18 months ago, Charles Kao, vice-chancellor of Chinese University, who did pioneering research in fibre optics in the 1960s, decided to do something about the problem. He persuaded six institutions — Chinese University, the University of Hong Kong, the Hong Kong University of Science and Technology, which will open in October, the Hong Kong Baptist College, the Hong Kong

Polytechnic and City Polytechnic of Hong Kong — to work together on a report that would highlight technological opportunities for Hong Kong.

Helped by a grant of HK\$200,000 (US\$25,000) from the Department of Industry, 60-70 professors and lecturers from the six institutions described areas of technology that have particular potential for Hong Kong. The report, entitled *Technology Road Maps for Hong Kong: An In-depth Study of Four Technology Areas*, was published last month.

The report pinpoints information technology, biotechnology, material technology and environmental technology as four fields that can make the most of Hong Kong's advantages. These include, the report says, excellent telecommunications, a strong entrepreneurial attitude, access to basic research and low-cost labour over the border in China, the availability of technological manpower, access to advanced technology from around the world and a keen perception of worldwide market demands and sentiments.

Projects recommended in the report can be undertaken with relatively small amounts of capital investment.

"Since many such opportunities concern specialized niche products, they may not be readily recognized by businessmen and entrepreneurs," the report says. "The experience of the present study shows that the tertiary education sector is able and ready to deliver technical assistance in identifying and evaluating these potential opportunities."

The scheme offers benefits for academics as well as industrials. At present, explained Robert Wu, director of the Office of Industrial and Business Development at the Chinese University of Hong Kong, the 3,000 - 4,000 PhDs who live in Hong Kong are largely underused. The six institutions concentrate largely on teaching, and offer few opportunities for research.

The collaborative effort already has one project in the works. It is putting together a team that will bid on a government project in environmental technology — one of the four areas spotlighted by the Road Map report. The Hong Kong government wants to identify the effect of its environmental regulations — many of which are not fully enforced — on Hong Kong's manufacturers. Such government contracts, worth many thousands of dollars, normally go to companies based outside Hong Kong.

Further up the road, it is likely that Varsity Resources will help companies to move into fields highlighted by the technology road maps report. Yiu-chung Cheng, director of City Polytechnic, re-

ported last week that the group has already received preliminary enquiries from industry.

One academic/industrial partnership in Hong Kong is already flourishing. Last week, computer company Unisys (Hong Kong) Ltd and the City Polytechnic of Hong Kong celebrated the first anniversary of a research and development agreement to develop computing products in various Asian languages by renewing their contract, this time for two years.

The enterprise started with two researchers from City Polytechnic and two from Unisys. Today, the City Polytechnic contingent has risen to ten, while eight Unisys staff members are working on the cooperative project. Later this month, Unisys plans to announce a series of products that the partnership has produced.

So effectively has the project helped the company to develop new products and skills "that I'm surprised that no one in Hong Kong has established similar arrangements," said Roy Clements, general manager of Unisys (Hong Kong). "Our competitors are missing a great opportunity to establish an important market niche."

Peter Gwynne

INDIA

Evolution on the air

New Delhi

ON 2 JUNE, the state owned All-India Radio (AIR) embarked on a massive science popularization programme — a 130-part serial on human evolution. Aired every Sunday in 17 languages from 51 radio stations, the serial will run for two-and-a-half years.

In all, 10,000 schools and 100,000 children from 10 to 14 years old are 'registered' listeners. They were signed up before the programme was launched and given a set of supplementary printed material and 'activity kits' — one set for each child and five sets for each school — designed to supplement the knowledge gained from the broadcasts.

Listeners can mail their queries to the nearest radio station and have them answered on later shows. One transmission after every three broadcasts is devoted to answering questions and after every 20 episodes listeners will participate in a quiz competition, with winners visiting Delhi to receive their prizes.

The Madras station alone has received 2,500 queries, mostly from children in rural areas without television.

The first prime minister of India, Jawaharlal Nehru, blamed India's backwardness on lack of scientific knowledge among the masses and urged scientists to teach them the scientific method. The radio serial, which will end in December of 1993, is a step in that direction.

K. S. Jayaraman

Growth of private funding

SIR — I should like to respond to the five responses to my letter (*Nature* 350, 370; 1991) that you have published.

D. H. Roberts, the provost of University College London (UCL), complains of inadequate government funding for science (*Nature* 350, 550; 1991). But, from the window of his own office, he can see the site for Easai's new UCL research institute. This institute, which will enjoy academic freedom, will also enjoy £50 million from the private sector. This illustrates my point exactly: British science is growing, but its growth is privately funded.

Frank Ashall and Alison M. Goate (*Nature* 350, 550; 1991) publish a graph showing that, over the past five years, authors in Britain have published as many papers in *Nature* as those in Germany, France and Japan combined. Why is this a source of complaint?

Peter Collins, with characteristic courtesy, makes two points (*Nature* 351, 9, 1991). First, he claims that the increase in the journal coverage by the Institute for Scientific Information (ISI) between 1973 and 1982 reflects the judgements of ISI, and not necessarily the growth of science. Yet the bibliometric and citation criteria that ISI employs makes it clear that its judgements are largely determined by the growth of science (E. Garfield, *Current Contents*, 28 May 1990, 5–13). Second, Collins is worried that so many short-term academic staff are of technical, not scientific, status. Yet surveys by the Science and Engineering Research Council (SERC) show that Britain enjoys a respectable number of principal investigators. Our major manpower shortfall lies in technical and support staff (*Research in the UK, France and West Germany: A comparison*, SERC, 1990). Fortunately, the free market is providing them.

Colin Humphreys (*Nature* 351, 513; 1991) claims that I “argue that British science is growing, relative to the other major industrialized countries”. This is not so. I argue that British science is growing absolutely (the facts are undeniable). However, British science grows less quickly than that of other industrialized nations because, unfortunately, Britain's economy grows so slowly (A. Maddison, *Phases of Capitalist Development*, OUP, 1982). Since international comparisons show that the major determinant of science funding appears to be gross domestic product (GDP) per capita (B. R. Williams, *Minerva* 3, 57–71; 1964) it follows that our science will experience relative decline (T. Kealey, *Scientometrics* 20, 369–394; 1991). But relative decline is not absolute decline. We grow.

John Mulvey (*Nature* 351, 513; 1991) complains that Britain spends too little on science as a percentage of GDP. But it has long been known that the percentage of GDP spent on research and development increases with wealth (B. R. Williams, *Minerva* 3, 57–71; 1964). Japan and Switzerland spend nearly 3 per cent (almost all of it privately funded), Malawi less than 0.03 per cent and Britain somewhere in the middle as befits our economic status. Mulvey's second point on our lack of industrial preparedness in 1914 is economically incorrect. In 1914 we were 25 per cent richer than Germany in terms of GDP per capita (P. Bairoch, *J. Eur. Econ. Hist.* 5, 273–340; 1976) and 27 per cent more industrialized (P. Bairoch, *J. Eur. Econ. Hist.* 11, 269–333; 1982). Undeniably, we were unprepared in 1914 in certain areas (notably khaki dye for military uniforms) but that followed on our relatively small expenditure on defence before 1914. Is Mulvey now arguing for more defence expenditure?

TERENCE KEALEY

University of Cambridge,
Department of Clinical Biochemistry,
Addenbrooke's Hospital,
Hills Road, Cambridge CB2 2QR, UK

Animal rights

SIR — You claim (*Nature*, 351, 517; 1991) that whereas the arguments propounded by the animal rights movement “go for the heart”, those of the biomedical community “go for the head”. A sociologist might suspect that the vital organ for vivisectionists is in fact the wallet. (No other members of the scientific community are nearly as incensed by the animal rights movement as those whose careers it threatens.) In any case, it is easy to demonstrate that under their pretence of intellectual objectivity, vivisectionists actually rely on nothing more rational than the age-old appeal of prejudice, of love of one's own kind.

Any vivisectionist would react with horror to the idea of experimenting on brain-damaged humans, or on human infants. Yet chimpanzees are fair game. This simple observation shows that vivisectionists are not primarily concerned with quality of mental function in deciding whether or not to slice up a given organism. There is no room to go into detail here, but such arguments-by-observation can be used to exclude all phenotypic characteristics as essentially irrelevant to the vivisectionist. For vivisectionists, the primary determinant of the moral rights of an organism appears to be whether or not the organism has human DNA. No characteristic

of the genome's phenotypic expression is considered nearly as important as the genome itself.

Vivisectionists apparently view this reverence for human DNA as axiomatic, in the sense that it is never justified as the conclusion of a line of reasoning but is rather assumed *a priori*. Surely even apologists for the biomedical community can see that such a morality has no more to do with “dry, reasoned discourse” than a morality valuing Caucasian DNA over non-Caucasian DNA, or a morality valuing male DNA over female DNA. The difference between vivisectionists and racial bigots is a matter of degree, not kind. Vivisectionist simply draw their lines in the sand at a further remove from themselves than racial bigots do. You do your readers a disservice by claiming that such arbitrary line-drawing reflects some sort of “intellectual rigour”.

DAVID WOLPERT

Theoretical Division and Center
for Non-Linear Studies, LANL,
Los Alamos, New Mexico 87545, USA

Bad taste

SIR — John Parks' letter (*Nature* 351, 434; 1991) regretting the fashion-driven perpetuation of methodological errors in animal feeding experiments reminds us of similar errors in diabetes research.

Tissue alterations reminiscent of those found in the complications of diabetes can be produced in many animals by the feeding of a diet consisting of 10 to 25 per cent galactose. Experimental drugs designed to mitigate the pathological consequences of galactosaemia are then added as dietary supplements. However, a common (but rarely observed and much less reported) effect is that the drugs taste noxious, so the animals feed less, consume less galactose and are thus “protected” against tissue damage relative to the control groups (see, for example, A. C. S. Woollard, Z. A. Bascall, G. R. Armstrong & S. P. Wolff, *Diabetes* 39, 1347–1352; 1990).

Unscheduled dietary restriction also causes unrecognized reduction of plasma glucose levels in true experimental diabetes and is similarly notorious in the examination of dietary supplements supposed to inhibit age-related pathology. It is not necessarily the case that Kantor's law of the perpetuation of ignorance is the problem. Rather, feeding experiments look simple but are so complex in design, execution and interpretation that they are rarely of value.

SIMON P. WOLFF

ALISON C. S. WOOLLARD

Department of Clinical Pharmacology,
UCMSM, University College London,
5 University Street
London WC1E 6JJ, UK

The fangs of the VIPER

Donald MacKenzie

VIPER is the first commercially available microprocessor whose design is claimed to have been proven correct. The controversy provoked by the claim reveals fundamental disagreement about the meaning of 'proof'.

COMPUTER systems are increasingly taking on roles where their failure could have catastrophic results, for example in medical care and in the control systems of aircraft and nuclear power stations. How can such systems be known to be safe? Certainly, they can be tested. But for a system of any complexity, the number of all possible combinations of external inputs and internal states is too large for even the most highly automated testing to be comprehensive. Nor does it necessarily help to install systems in triplicate, as is often done with crucial electromechanical systems. This is a good insurance against physical failure, but not against a hardware or software design flaw common to all three systems.

Because of this, computer scientists have been seeking ways to prove mathematically that the design of computer systems is correct. In 1986 the UK Cabinet Office's Advisory Council for Applied Research and Development suggested that such mathematical proof should eventually become mandatory for systems where failure could involve more than ten deaths¹.

A major step in this direction came in the late 1980s when a team of researchers employed by the UK Ministry of Defence developed a novel microprocessor called VIPER (verifiable integrated processor for enhanced reliability). Though VIPER has several other features designed to make it safe (such as stopping if it encounters an error state), what is crucial about it is the claimed existence of a mathematical proof of the correctness of its design — something no other commercially available microprocessor can boast.

Controversy

The claim has, however, been controversial. There has been sharp disagreement over whether the chain of reasoning connecting VIPER's design to its specification can legitimately be called a 'proof'. Lawyers and judges, as well as academics, became involved. Charter Technologies, a small English firm which licensed aspects of VIPER technology from the Ministry of Defence, took legal action against the ministry in the High Court, alleging that the claim of proof was a negligent misrepresentation. The ministry vigorously contested this allegation, and Charter went into liquidation before the case could come to court. If it

had, the court would have been asked to rule on what constitutes mathematical proof, at least in the context of computer systems. Lawyers and judges would have had to grapple with esoteric issues previously the province of mathematicians and logicians.

The home of VIPER is the Ministry of Defence's Royal Signals and Radar Establishment (RSRE), the leading research and development laboratory in Britain in radar, semiconductor physics and several fields of information technology. VIPER was developed by a team of three at RSRE, led by John Cullyer, a man concerned by the potential for computer-induced catastrophe^{2,3}. In the VIPER project, Cullyer and his colleagues, John Kershaw and Clive Pygott, aimed to design a microprocessor whose detailed, logic-gate level design could be proven to be a correct implementation of a top-level specification of its intended behaviour. With the methods available in the 1980s, a direct proof of correspondence was out of the question, even for the relatively simple VIPER chip. Instead, the team sought to construct the proof in the form of a chain of reasoning connecting four levels of decreasing abstraction.

Most abstract is the top-level specification (level A), which lays down the changes that should result from each of the limited set of instructions provided for use by VIPER's programmers. Level B, the major-state machine, is still an abstract description, but contains more detail on the steps gone through in executing an instruction. Level C, the block model, is more concrete, and consists of a diagram (of a kind familiar to integrated circuit designers) of the major components of VIPER, together with a specification of the intended behaviour of each component. In level D the description was sufficiently detailed to be usable to control the automated equipment employed to construct the 'masks' needed to fabricate VIPER chips.

At the end of 1986, the RSRE team began its authoritative account: "VIPER is a 32-bit microprocessor invented at RSRE for use in highly safety-critical military and civil systems. To satisfy certification authorities of the correctness of the processor's implementation, formal mathematical methods have been used both to specify the overall behaviour of the processor and to prove

that gate-level realisations conform to this top-level specification"⁴. But the details of the paper made clear that this was as much a statement of intention as of achievement. Crucial links in the chain of proof were still being forged. Before this work was complete, however, VIPER began to be sold, both literally and metaphorically. The two processes were to collide disastrously.

Market forces

During 1987 and 1988, with defence research establishments like RSRE under considerable political pressure to show commercial benefit from their work, VIPER was moved rapidly onto the market. Ferranti and Marconi undertook to make prototype chips, using different processes as an insurance against errors being introduced in their physical production. Several firms took up the equally essential work of making it possible for users to write programs for VIPER and to check their correctness. Most central was Charter Technologies, which in 1988 signed a contract with the Ministry of Defence to sell and support the VIPER software toolset developed by RSRE. Through the intermediary of Defence Technology Enterprises, Charter also purchased the licence to develop and sell the VIPER prototyping system. The company marketed VIPER actively⁵. The technical press was unrestrained in its welcome, claiming that VIPER had been mathematically proven to be free of design faults⁶⁻⁸.

Commercialization and glowing media accounts of VIPER had a double-edged effect. In 1986, RSRE had awarded a contract to the University of Cambridge, to investigate the possibility of mechanically checking the higher-level VIPER correctness proofs. Cullyer had used an automated system, LCF-LSM, developed at Cambridge, to write VIPER's specification and to construct an outline 'paper-and-pencil' proof that the major state machine (level B) was a correct implementation of it. By January 1987, Avra Cohn at Cambridge successfully used HOL (ref. 9), LCF-LSM's successor, to formalize Cullyer's outline proof linking levels A and B. The pencil-and-paper proof took Cullyer three weeks. The mechanized proof, in which every step of logical inference is made explicit, took six person-months to set up, and

consisted of a sequence of more than a million inferences¹⁰.

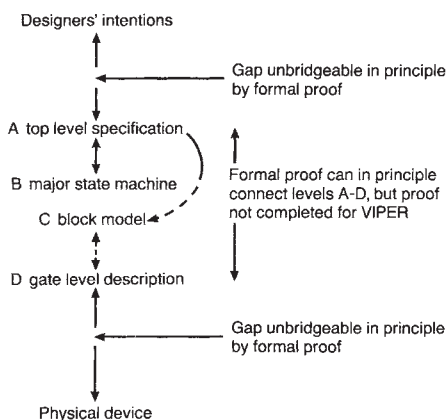
Cohn then embarked on the more ambitious task of proving that the block model (level C) was a correct implementation of the top-level specification. She constructed a partial proof of more than 7 million inference steps, but by May 1988 concluded that to complete the proof would mean a great deal more work for what she called a "dwindling research interest", as well as requiring further development of the HOL system. Cohn had also become worried about the way VIPER was being described in marketing material and in the media. The very language of proof and verification could, she felt, convey a false sense of security. To say that the design of a device was 'correct' does not imply that the device would do what its designers intended: it means only that it is a correct implementation of the formal specification. Even the most detailed description of a device was still an abstraction from the physical object. Gaps unbridgeable by formal logic and mathematical proof must remain between VIPER's level A description and the designers' intentions, and between level D and the actual chip (see figure)¹¹.

VIPER's developers were perfectly aware of these unavoidable limitations of formal proof. The gap between gate-level descriptions and physical chips was indeed the reason they had sought independent physical implementations of VIPER, implementations that can (in the 1A version) be run in parallel, checking each other. What does seem to have been more of a shock to them, however, was a highly critical assessment from the US verification specialists, Computational Logic, Inc., of Austin, Texas. The firm had been commissioned by NASA to assess the VIPER proof. The report's authors, Bishop Brock and Warren Hunt, were prepared to grant the status of formal proof only to Cohn's HOL proof linking levels A and B. All other links in the chain were, in their view, either incomplete or informal. For example, the RSRE team had initially established the correspondence levels C and D by simulation and case-by-case checking, using a method they called "intelligent exhaustion"¹². The simulation, however, required LCF-LSM specifications to be translated into the RSRE-developed ELLA language, and there was no formal proof of the correctness of the translation. Nor, said Brock and Hunt, was there a proof that all possible cases had been considered. VIPER, they concluded, had been "intensively simulated and informally checked", but not "formally verified"¹³.

News of the impending report reached the VIPER team in September 1989, when Charter's managing director,

Digby Dyke, met Cullyer at a seminar. According to Dyke, Cullyer told him that the report was "damning" and would "kill VIPER dead". When the RSRE team finally received a copy of the draft, it appeared that Cullyer conceded Brock and Hunt's criticisms. The version it was sent by fax from Austin had been shown previously to Cullyer, and contained his handwritten annotations and signature. The word 'agreed' appeared repeatedly in the margin.

Cullyer's response was generous, especially given that Computational Logic could be seen as a rival, since the firm was developing a formally proven microprocessor of its own. The RSRE team was happy to acknowledge that "more remains to be done, both to build up confidence in the existing VIPER design and to develop new techniques of design



and verification which avoid the limitations of present methods"¹³.

But by autumn 1989, VIPER was no longer just a research project. It was a commercial product, and one that was not meeting with great success, quite apart from any problems of the proof. Potential users were reluctant to abandon tried and trusted microprocessors (even with their known bugs) for a novel chip and associated new software, and VIPER was too simple and slow for many applications. Government policy prohibited the Ministry of Defence from mandating the use of VIPER, and only one defence project adopted it. The only civil adoption was for a system to control signals on automatic crossings on the trans- and central Australian rail routes.

The combination of commercial failure and the criticism of the claim of proof broke apart the previously close ties between RSRE and Charter. Charter began by seeking informal redress for the losses it believed it had suffered, then took its grievances to members of parliament, the media and (by January 1991) to the law courts.

In the ensuing storm of publicity concerning the dispute, the central issue raised by the episode has frequently been submerged. No 'bug' has been

found in VIPER chips: indeed, their design has been subjected to an unprecedented degree of scrutiny, checking, simulation and mathematical analysis. The key question, rather, is whether the results of this process amount to a 'proof'. As we have seen, to Brock and Hunt they did not. Cohn, similarly, wrote in 1989: "no formal proofs of Viper (to the author's knowledge) have thus far been obtained at or near the gate level"¹⁴. Others, however, would note that most of the mathematics has not been proven in this fully formal sense, and would claim that it is an unduly restrictive notion of 'proof'.

In the end, the High Court was not called on to take sides in this ultimately philosophical dispute. Yet the Ministry of Defence in April 1991 issued the standard "The Procurement of Safety Critical Software in Defence Equipment"¹⁵. Although there are many other measures proposed, and provision for exceptions, there is a clear pointer to the desirability of fully formal proof in the most crucial applications.

As the ministry, and other major users of safety-critical computer systems, move in this (wholly praiseworthy) direction, it is important that the lessons of the past be learnt rather than suppressed. The VIPER episode reminds us that 'proof' is both a seductive word and a dangerous one. We need a better understanding of what might be called the 'sociology of proof': of what kinds of argument count, for whom, under what circumstances, as proofs. Without such an understanding, the move of 'proof' from the world of mathematicians and logicians into that of safety-critical computer systems will surely end up, as VIPER nearly did, in the courts. □

Donald MacKenzie is in the Department of Sociology, University of Edinburgh, 18 Buccleuch Place, Edinburgh EH8 9LN, UK.

1. Advisory Council for Applied Research and Development, *Software: A Vital Key to UK Competitiveness* (HMSO, 1986).
2. *Computer Weekly* (4 February 1988).
3. *The Times* (29 January 1988).
4. Cullyer, W. J. & Pygott, C. H. *IEE Proc.* **134E**, 133-141 (1987).
5. Charter Technologies Ltd *VIPER Microprocessor Development Tools* (December 1987).
6. *New Scientist* (16 October 1986).
7. *Electronics Weekly* (15 October 1988).
8. *The Engineer* (4 February 1988).
9. Church, A. J. *symb. Logic* **5**, 56-68 (1940).
10. Cohn, A. *A Proof of Correctness of the Viper Microprocessor: The First Level* (University of Cambridge Computer Laboratory Technical Report No. 104, 1987).
11. Cohn, A. *Correctness Properties of the Viper Block Model: The Second Level* (Univ. Cambridge Computer Laboratory Technical Report No. 134, 1988).
12. Pygott, C. H. *Formal Proof of Correspondence between the Specification of a Hardware Module and its Gate Level Implementation* (RSRE Report 85012, 1985).
13. Brock, B. & Hunt, W. A. Jr. *Report on the Formal Specification and Partial Verification of the VIPER Microprocessor* (Computational Logic Technical Report 46, 1990).
14. Cohn, A. J. *autom. Reasoning* **5**, 127-139 (1989).
15. Ministry of Defence *Interim Defence Standard 00-55 (Part 1/Issue 1)* (5 April 1991).

Towards the brain-computer's code?

A simple model seems to suggest how information about an external stimulus may be gleaned from the trains of pulses neurons of the sensory systems send to the central nervous system.

ALTHOUGH it is generally accepted that the brain is a computer of a kind, nobody yet knows what is the computer code or codes (or there may be different codes for different purposes). The idea that memory may be encoded in the sequence of the chemical subunits of macromolecules is natural enough — there is so much memory to accommodate — but the simplest versions of it current only a few years ago cannot easily be reconciled with the interconnectedness of the neurons of the cerebral computer, or with the likelihood that there is something in the neural network view of how the brain works.

But it is also plain that such a code could not possibly apply to the way in which information is transmitted from the periphery to the central processing regions, where it seems now to be well-established that the activity of a neuron, like the information traffic along a neuronal channel, is measured best by the rate at which voltage pulses, or spikes, are generated within it. In due course, it will be necessary for people to be asking themselves how two such radically different coding principles are mated with each other.

Indeed, the only coded information in the brain that is, for the time being, vaguely susceptible to measurement is that from the neurons of the peripheral sensory system, but even then the meaning is far from clear. Does the rate at which a peripheral neuron "fires" provide a simple measure of the intensity of the stimulus, or of something more subtle? How is phase information embodied in the signal and — if it is not — how else can the signals from neighbouring neurons in the same sensory system be combined with each other to yield other than averages? And is the noisiness of the output of the sensory neurons simply a nuisance, or may it be more significant?

This, more or less, is the starting point for an intriguing argument about the mechanism of sensory coding in the central nervous system by André Longtin from the theoretical division at the Los Alamos National Laboratory, Adi Bulsara from the Naval Systems Center at San Diego and Frank Moss from the University of Missouri at St Louis. Interest apart, one reason for drawing attention to the article is that it appears in what must be, for most neurophysio-

logists, an obscure journal: *Physical Review Letters* (67, 656; 1991).

The starting-point is experimental — measurements of the intervals of time between successive spikes from sensory neurons stimulated by a periodic sensation. In the neurophysiology trade, it is standard practice to measure the intervals of time between successive spikes of voltage, and then to plot these as a histogram. If the stimulus is large enough, the result is much as would be expected — a peak centred on the time interval corresponding to the period of the exciting stimulus and then successively smaller peaks centred on time-intervals that are integral multiples of the underlying period. To put the point almost anthropomorphically, it is as if the excited neuron does its best to respond at intervals determined by the period of the exciting stimulus, but that it sometimes fails to do so, whereupon it will try its best to do so when the stimulus is next a maximum and, if not then, then on the next occasion and so on . . .

Luckily, there is a little more than that to say, which is what Longtin and his colleagues attempt. They begin with data gathered more than 20 years ago from single auditory nerve fibres of the monkey and from single neurons in the primary visual cortex of the cat, each of them exposed at the periphery to a periodic stimulus. The histograms of the intervals between successive voltage spikes peak at the period of the stimulus and at integral multiples thereof. It is crucial to the argument that the intensity of the successive peaks decays exponentially, suggesting that the underlying process is stochastic in some sense.

The objective is to "construct the simplest possible physical mechanism" that captures the behaviour of the histograms. That is laudable enough, though whether it can be fairly described, as the authors describe it, as taking "a reductionist's view" of the data is another matter. Perhaps the time has come to distinguish between those who believe that minimalist models have inherent virtues and those who hold that correct models will be found to be minimalist; they might be called reductionists of the first and second kinds respectively.

That quibble apart, the minimalist model is as simple as anybody could ask. If a neuron is a two-state system, say ON or OFF, most conveniently corresponding

to the condition in which the transmembrane potential is sufficient to allow the neuron to fire and that in which the membrane is hyperpolarized, as it is after an action potential has been triggered, then why not represent that two-state system by a mechanical system governed by a double-well potential? Then the question to be decided is the rate at which the mechanical system will make its way from one well to the other, which will be determined by the external stimulus (say a sinusoidal force with fixed period), the noise and the shape of the potential (notably the height of the barrier between the two wells).

Now, in principle, it is possible to calculate the behaviour of the simple model. The underlying model for the neuron shows it making transitions from one state to another, but to what do the time intervals between successive spikes of potential correspond? One possibility is that the intervals of time actually measured represent the times the system spends in one state or the other (because, since the problem is made symmetrical for simplicity, either one or the other state will give the same result). Or the measured time intervals may represent the times between successive transitions from one particular state to the other. As it happens, the mathematics yields a straightforward answer. The first interpretation should yield a histogram with peaks at every odd-integer multiple of half the underlying period, the second at integral multiples of the underlying period. It follows from that that the model entails the assumption of a "recovery" phase of some kind after the discharge of each voltage spike.

The other general finding is that of the significance of noise. If the model is anything like correct, the root mean square noise potential and the external stimulus add together arithmetically in determining the intervals between the firing of a neuron. What that means is that neuronal noise may be a way of magnifying the signal that must be interpreted with the help of the code that applies elsewhere in the nervous system. Much depends, of course, on the exact nature of the code employed in those central regions, but this could begin to qualify the general assumption that the ratio of signal to noise is a universal figure of merit for the behaviour of a communications system. **John Maddox**

Aerodynamics in a spin

S. H. Salter

MECHANICAL engineers with a bent towards innovation should be on the alert for reports on unusual fluid dynamical behaviour. The paper on autorotation by B. W. Skews on page 512 of this issue is a good example. This describes an experimental study on the aerodynamics of various multifaceted shapes rotating in an airflow and gives findings which could be applied in fields as diverse as aircraft accident investigation and nuclear safety.

Skews tested regular prismatic polygons which were free to rotate about their long central axis in a wind tunnel with the air stream perpendicular to the axis of rotation. He measured lift, drag and rotation speed for various air speeds on polygons with two to eight sides at Reynolds numbers ranging from 50,000 to 200,000. Skews finds the ratio of polygon edge-speed to wind-speed to be a significant parameter that is nearly independent of the enclosing diameter of the polygon. Triangular sections give the highest ratio at about 0.6, but square sections spin as fast as flat plates with edge-speed to wind-speed ratios of 0.45.

The lift coefficients are quite respectable (up to 1.4 for a spinning flat plate). The drag coefficients fall slightly with increasing number of sides but are not much above the 1.2 which would be expected for a cylinder at these Reynolds numbers. A slightly ghoulish result is that the lift/drag ratios allow an estimation of the free glide angles for broken pieces of aircraft. The angle would be about 45° for objects like a wing. By measuring the spread of debris we can tell the height of the break-up.

The innovative engineer cannot help looking for means of exploiting such new observations. Efforts fall into two distinct categories: those in which the phenomena might be useful and those in which their influence might be detrimental. Unfortunately, the useful category for Skews' work does not seem to be very promising. The edge-speed ratios are much lower than the range 5–10 which are usual for the tip speeds of wind turbine rotors and they would be lower still if the polygon was providing any driving torque. The S-shaped Savonius rotor is already doing a better job and even that is out of favour for energy generation. A set of Skews' triangular prisms on a squirrel-cage rotor could make a formidable egg-beater; chemical engineers wanting to stir large vats of liquid using vortex injection, take note.

The detrimental aspects seem to be the more important outcome of Skews' work. Multistrand ropes and cables have a quasi-polygonal shape and long ones

are fairly free in rotation. Skews may help us understand their behaviour in cross winds. An even more exciting area might be the fluid dynamics of nuclear reactors. Nuclear engineers are fond of hexagonal packing of fuel rods. They also like long rods, which are inevitably

the cylinder shape allows a much stronger internal spar than a conventional aerofoil and behaviour is less affected by changing angle of incidence. The addition of an anti-flow-separator plate by J. Cousteau and Pechiney in 1985 (B. Charrier *et al.* in *Windship Technology; Proc. Int. Symp. Windtech '85* (ed. C. J. Satchwell) 39–61 (Elsevier, Oxford, 1985)) led to further improvements in performance. Flettner was unlucky to do his work just before a major economic

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The first rotor ship, the Buckau, was a converted schooner. With a full load, the ship could reach a speed of 14.3 km hr⁻¹ driven only by the rotors. With counter-rotating rotors, the ship could turn on the spot. It entered transatlantic service in 1926.

compliant in rotation and deflection and which have to be placed in high-velocity coolant flows. Mixing rotation and deflection does them no good at all. Perhaps Skews' observations will prevent a nuclear meltdown.

The innovative engineer should also build mental pathways to similar and dissimilar ideas. Skews mentions the Magnus effect exploited in the Flettner rotors which were used successfully during the late 1920s to drive several sailing ships (see figure) and were even used for Atlantic crossings. Flettner used vertical rotors which were only about twice the height and diameter then favoured for ship funnels. They could be rotated about a vertical axis at speeds up to a few revolutions per second. Force was generated at right angles to the wind, proportional in strength to the product of wind speed and rotor circulation.

The forces would correspond to very high lift coefficients (up to 9) compared with a conventional aerofoil. This meant that the rotor area would be much lower than the equivalent sail area. Drag coefficients as low as 0.15 are reported by F. Diaz *et al.* (*J. Am. Inst. Aeronaut. Astronaut.* 23, 49–54; 1985), higher than for an unstalled aerofoil but lower than for a non-rotating cylinder. Furthermore

depression and a period of falling bunker fuel prices. Some engineers retain an affectionate interest in his ideas and hope that their time may come again.

Perhaps the most intriguing mental pathway leads to the little-known Budal effect observed when a horizontal cylinder spins in beam waves. Kjell Budal was one of the first Norwegian wave-energy pioneers, working with Falnes at Trondheim. Among his many ideas was an apparently magic way to magnify or reduce the effective size of a marine structure. This would have the effect both of reducing the quantity of material needed to build a wave-energy converter and also reducing the extreme loads.

It is well known that the pattern of water movement in a deep-water travelling wave is a circular orbit with diameter at the surface equal to the wave height. We can regard this movement as a rotating velocity field which is 90° in advance of a rotating acceleration field. For objects like mooring ropes which are very small in comparison to the wave amplitude, the forces felt depend on the square of local water velocity. But for bigger objects with sizes that are comparable to the wave amplitude, it is the acceleration field that dominates fluid loading. Force is needed to oppose the

acceleration which the water would have had if the body was not present. It is proportional to the wave height.

Budal realized that if we fit a spinning sleeve round a submerged horizontal cylinder we will be able to produce a lift as predicted by Magnus (K. Budal & P. M. Lillebeken *IUTAM Symp. Hydrodynamics Ocean Wave Energy* (eds D. V. Evans & A. F. de O. Falcao) 205–215 (Springer, Berlin, 1985)). This lift will be, by definition, 90° ahead of the velocity field and so will be parallel or antiparallel to the acceleration field. It follows that the lift will add to or, by changing spin direction, subtract from or even reverse the wave loads. Budal's work with small scale models has been

replicated. It is not yet known whether the effect will scale up as well as the Flettner rotor did. If it does, the implications for wave energy are profound. A 5-metre Evans' cylinder could behave like a 15-metre one but could progressively shed its wave loading when things got uncomfortable. All that it would need would be a thin rotating outer sleeve which would use about 1 per cent of the output power to drive it. Scaling-up the Budal effect may very well be related to the vortex shedding which makes the Skews' polygons rotate. □

S. H. Salter is in the Department of Mechanical Engineering, University of Edinburgh, Edinburgh EH9 3JL, UK.

CELL BIOLOGY

ABG of microtubule assembly

Caroline E. Alfa and Jeremy S. Hyams

RESOLUTION of the long-standing uncertainty over how the polar assembly of microtubules in living cells is organized has come a step nearer with the publication of two papers in *Cell*^{1,2}. Zheng *et al.*¹ and Stearns *et al.*² have discovered that the newest member of the tubulin superfamily, γ -tubulin, is specifically localized to the centrosome, the animal cell's microtubule organizing centre, and is thereby centrally implicated in the establishment of polarity.

In the relatively short period since its discovery³ γ -tubulin has intrigued microtubule researchers, who for 20 years previously had been content in the knowledge that microtubules were composed of α/β tubulin dimers^{4,5}. Although it has been shown through molecular techniques that the genomes of most organisms contain several α - and β -tubulin genes (vertebrates, for instance, contain six or seven of each⁶), the appearance of a completely new addition to the family was a great surprise. The discovery of γ -tubulin derived, not like that of α and β from classical grind-and-find biochemistry, but from a painstaking combination of classical and molecular genetics in that much underrated organism, the bread mould *Aspergillus nidulans*.

The gene *mipA* (for microtubule-interacting protein) was isolated as an allele-specific, extragenic suppressor of *benA33*, a heat-sensitive β -tubulin mutant of *Aspergillus* that at the restrictive temperature is defective in both mitosis and the movement of nuclei along hyphae (another microtubule-dependent process)⁷. When cloned and sequenced, *mipA* was found to encode a protein of 454 amino acids that is distinct from the α - and β -tubulins but about as similar to each as they are to one

another (just over 30 per cent similarity in amino acid sequence)³. Disruption of *mipA* resulted in spores that could germinate and undergo DNA synthesis, but that could not carry out nuclear division and therefore died⁸. Most surprisingly, an antibody to γ -tubulin failed to stain microtubules by indirect immunofluorescence microscopy but reacted specifically with the spindle pole bodies, the microtubule organizing centres of this and other fungi (see our previous News and Views article⁹).

These initial studies did not however address the key question — is γ -tubulin merely a curiosity of the *Aspergillus* microtubule cytoskeleton, or is it of more general significance? This issue has been unequivocally resolved in the *Cell* papers^{1,2} and another now in press¹⁰. Extending their own earlier studies, Zheng *et al.*¹ have now isolated γ -tubulin genes from *Drosophila* and *Homo sapiens*. The sequences of both are very similar to that of the original *Aspergillus* gene, meaning that γ -tubulin is likely to be a highly conserved protein. Moreover, antibodies to the fungal protein stained the centrosomes of both insect and human cultured cell lines, suggesting that γ -tubulin might be a ubiquitous component of microtubule organizing centres.

Centrosomes consist of two distinct elements, the centrioles and their surrounding shell of ill-defined pericentriolar material. Microtubules seed from the latter, and using immunogold electron microscopy, Stearns *et al.*² have demonstrated that it is this material with which γ -tubulin is associated. These workers have also greatly extended the survey of γ -tubulin gene distribution by isolating homologues from *Xenopus* and fission yeast. Again, both are closely related to

mipA. Moreover the authors report the isolation of partial sequences from a higher plant, an alga and a budding yeast. The fission yeast gene has also been isolated by Horio *et al.*¹⁰ who further showed that, as in *Aspergillus*, γ -tubulin antibodies stained the spindle pole bodies and also a second pair of microtubule organizing centres that appear transiently at the equator of late mitotic cells in this yeast.

How do these observations bear on microtubule assembly? As a consequence of their formation from aligned asymmetric subunits, microtubules have an intrinsic polarity. This is reflected in their growth, the fast-growing (plus) end adding or losing subunits with much greater avidity than the other, slower-growing minus end. In the interphase animal cell, microtubules are assembled with a common polarity, the minus ends being embedded in pericentriolar material, the plus ends being free to grow and shrink and explore their cytoplasmic environment. This arrangement is of enormous significance, as it provides a directional framework for cytoplasmic motor proteins such as dynein and kinesin to orchestrate a bidirectional traffic flow of vesicles and organelles. At mitosis, too, the polar assembly of spindle microtubules from the now duplicated centrosomes is essential for the establishment of balanced forces operating on sister kinetochores.

Up to now, the way that this polarity was achieved was entirely unknown. The key is almost certainly γ -tubulin, which was originally identified on the basis of its interaction with β -tubulin. As Zheng *et al.*¹ point out, if β -tubulin interacts with the α subunit in the microtubule organizing centre, then the α subunit is always distal to its organizing centre and polarity is *ipso facto* established.

The only possible cloud on the horizon of this appealing model is that most interesting things in biology come in pairs. If α -tubulin has its β counterpart, can we expect that γ has its hitherto undiscovered δ ? □

Caroline E. Alfa and Jeremy S. Hyams are in the Department of Biology, University College London, Gower Street, London WC1E 6BT, UK.

1. Zheng, Y., Jung, M. K. & Oakley, B. R. *Cell* **65**, 817–823 (1991).
2. Stearns, T., Evans, L. & Kirschner, M. *Cell* **65**, 825–836 (1991).
3. Oakley, C. E. & Oakley, B. R. *Nature* **338**, 662–664 (1989).
4. Bryan, J. & Wilson, L. *Proc. natn. Acad. Sci. U.S.A.* **68**, 1762–1766 (1971).
5. Feit, H., Slusarek, L. & Shelanski, M. L. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2028–2031 (1971).
6. Cleveland, D. W. *J. Cell Biol.* **104**, 381–383 (1987).
7. Weil, C. F., Oakley, C. E. & Oakley, B. R. *Molec. Cell Biol.* **6**, 2963–2968 (1986).
8. Oakley, B. R., Oakley, C. E., Yoon, Y. & Yung, M. K. *Cell* **61**, 1289–1301 (1990).
9. Alfa, C. E. & Hyams, J. S. *Nature* **348**, 484 (1991).
10. Horio, T. *et al.* *J. Cell Sci.* (in the press).

An unconventional family tree

N. d'Ambrumenil

FAMILIES of unconventional superconductors, like Tolstoy's happy families, all resemble one another. At least this is the conclusion of a new study by Uemura *et al.*¹ of correlations between the magnetic penetration depths and the transition temperatures for the onset of superconductivity in these materials.

The pattern emerges if the superconducting transition temperature, T_c , is plotted as a function of $1/\lambda^2$ where λ is the London penetration depth (see figure). Unconventional superconductors are those not well described by the standard 'BCS' theory, which assumes phonon-mediated interactions between electrons. Uemura *et al.* report results for the high- T_c cuprate superconductors, the heavy-fermion compounds UPt_3 and UBe_{13} , the Chevrel phases² (molybdenum sulphides, named after their discoverer), $\text{Ba}_{1-x}\text{K}_x\text{BiO}_3$ (BKBO), and the organic quasi-two-dimensional superconductor $(\text{BEDT-TTF})_2\text{Cu}(\text{NCS})_2$. The one example left out of the list, the new³ buckminsterfullerene-based superconductor K_3C_{60} , has since been added (see next week's *Nature*⁴) and also fits the same pattern.

Besides T_c , the unconventional superconductors have a second energy scale, T_F , set by their electronic interactions. The penetration depth, λ , provides an estimate of this energy. Uemura *et al.* find that when they plot $\log T_c$ against $\log T_F$ the points lie more or less on a straight line. This suggests that the transition temperature and, by implication, related properties of the superconductors are functions of only the energy scale T_F and are insensitive to details of the electronic interactions or dimensionality (two or three) of the various systems involved.

The authors actually report measurements of the muon spin relaxation (μSR) rate, σ , rather than λ : μSR measurements are known to yield the most accurate values for the 'London' penetration depth for many superconductors. Polarized muons fired into a magnetized superconductor precess at different rates according to the strength of the local field so that the original polarization is lost. Their spin relaxation rate is proportional to λ^{-2} , with, for systems with short superconducting coherence lengths,

$$\sigma \sim \frac{1}{\lambda^2} \sim \frac{n_s}{m^*} \quad (1)$$

Here, n_s is the nominal density of superconducting charge carriers, and m^* the carriers' effective mass (shifted by inter-

actions from the normal value).

Values obtained from μSR for n_s/m^* can be used to estimate the electronic energy scale or, where it makes sense to call it such, the effective Fermi energy, T_F . In the systems that are quasi-two-dimensional (the high T_c and BEDT superconductors) the relation between T_F and density is

$$T_F = \frac{\hbar^2 \pi}{2} \frac{n_s}{m^*} \quad (2)$$

The effective two-dimensional carrier density is assumed to be the three-dimensional density multiplied by the interplanar spacing.

It is a remarkable piece of luck that equations (1) and (2) both involve the ratio n_s/m^* . Measurements of σ therefore provide a direct measure of the energy T_F so that there is no need to interpret other experimental quantities to unravel n_s and m^* . For the high- T_c superconductors, T_F has values of around 1,000 K; this is comparable with the energy of spin-spin interactions. And in the case of the BEDT organic superconductor T_F is roughly 200 K — exactly the temperature at which the system has a metal-insulator transition.

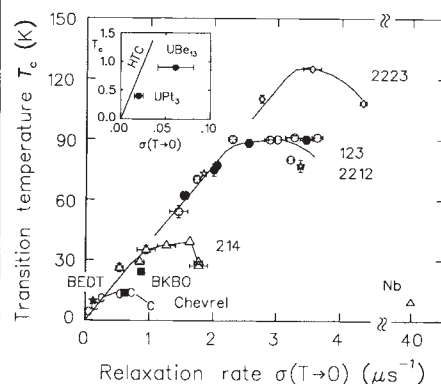
The low-energy electronic states in the organic BEDT superconductor are thought to be well-described by simple tight-binding 'Debye-Hückel' molecular-orbital type calculations. There are four simple energy bands near the Fermi energy, with only two, nearly degenerate, bands actually crossing the Fermi level⁵. Although strong electron-electron interactions have been supposed important in these compounds for some time (there is, after all, a metal-insulator transition at 200 K), the direct measurement of the effective bandwidth reported by Uemura *et al.* should stimulate more theoretical interest in the BEDT family. The members have all the ingredients theorists like: few orbitals per molecule, two-dimensionality and a direct measurement of the relevant energy scale.

For the three-dimensional compounds, in the corresponding relation to equation (2), T_F varies as $n_s^{2/3}/m^*$. This ratio is not measured by μSR and it is necessary to combine the results obtained for σ with estimates of m^* taken from specific-heat measurements, to obtain an estimate for T_F . This involves assuming that, at low temperatures, the systems are 'Fermi liquids' (in which the charge carriers behave as if they were independent) which, for the compounds involved (BKBO, the heavy-fermion compounds

and the Chevrel phases), is thought to be the case. The estimates obtained for T_F are of the same order of magnitude as the (somewhat ill-defined) spin-fluctuation temperature in the heavy-fermion compounds and only slightly larger than the metal-insulator and structural phase transition temperatures in BaKBiO_3 and in the Chevrel phases, respectively.

The energy T_F is an energy or temperature below which some degrees of freedom are effectively quenched. In the heavy-fermion and high- T_c superconductors these are spin degrees of freedom. In the other compounds it is certain charge configurations. Where it is the charge degrees of freedom which are excluded, there is usually a phase-transition associated with the energy scale T_F . This can be a structural phase transition as in the Chevrel phases, in which electrons are seen to localize in certain Mo d-orbitals^{2,6}, or a metal-insulator transition (BEDT), or both (BKBO).

For energies within T_F of the ground-state energy, the excitations of the system exist in the restricted space of the unquenched degrees of freedom. However, the large increase in entropy associated with the 'missing' degrees of



The superconducting families tree. The 'unconventional' superconducting families all lie on the same straight line in the above plot for small superconducting current carrier densities and branch off in similar ways once a critical density is reached. Actually plotted is the muon spin relaxation rate σ in the limit $T \rightarrow 0$; σ is proportional to the superconducting density, n_s , and to λ^{-2} , with λ the penetration depth (equation (1)). (The relaxation rates for the heavy-fermion compounds UBe_{13} and UPt_3 are not accessible to μSR , and have been converted from independent measurements of λ .) The numbers 2223, 123, 2212 and 214 are abbreviations for the high- T_c cuprate families based on: $\text{Ti}_2\text{Ba}_2\text{Ca}_2\text{Cu}_3\text{O}_{10}$, $\text{YBa}_2\text{Cu}_3\text{O}_7$, $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ and La_2CuO_4 . The Cs denote the results for the Chevrel phases; the solid squares, those for the BKBO family; and the solid star, that for BEDT. K_3C_{60} (not shown) has⁴ $T_c = 19.3$ K and $\sigma = 0.31 \mu\text{s}^{-1}$. The straight line in the inset corresponds to the linear relation found for the cuprates.

freedom, which must occur between absolute zero and T_F , implies a large influence on the parameters which describe the low-energy excitations. These include the effective mass if the system remains a Fermi liquid, the symmetry of the superconducting ground state, and, in the more exotic theories invoked to explain the high- T_c compounds, even the statistics. (Of course identifying the low-energy excitations, their quantum numbers and the effect of the virtual excitation of the excluded high-energy states is what everyone is trying to do.)

The measurements of λ can also be used to estimate a Bose-Einstein condensation temperature. If the superconducting carriers were all to bind to form local 'Bose' pairs (this is in contrast to conventional superconductors in which pairs extend over hundreds of lattice sites) and if these were assumed to be non-interacting then the systems would become superconducting at the Bose-Einstein condensation temperature, T_B . In reality the coherence length measured in these compounds suggests that if there were pairs they would overlap and interact strongly which in turn would reduce the transition temperature: T_B would then be an upper bound for the transition temperature. Uemura *et al.* find that, for the 'unconventional' superconductors they study, the observed transition temperatures are only a factor of 3–30 times smaller than T_B and between 10 and 100 times smaller than T_F . They also quote corresponding results for the conventional superconductors Nb, Sn, Al and Zn. There is no obvious pattern: Nb is far from the family tree (see figure) and all have T_c 10^4 – 10^5 times smaller than T_B .

That T_c is essentially proportional to T_F and T_B in the unconventional compounds is consistent with the 'strong-coupling' models — resonant valence bonds, anyons, negative-U Hubbard models — that have been proposed to explain their superconductivity. These assume unretarded pairing interactions between carriers and, at least conceptually, local pairs.

As yet unexplained remains the result that once above a certain critical density of carriers, n_c , the compounds all seem to branch off from the tree (see figure) in much the same way. □

N. d'Ambrumenil is in the Department of Physics, University of Warwick, Coventry CV4 7AL, UK.

Out of equilibrium

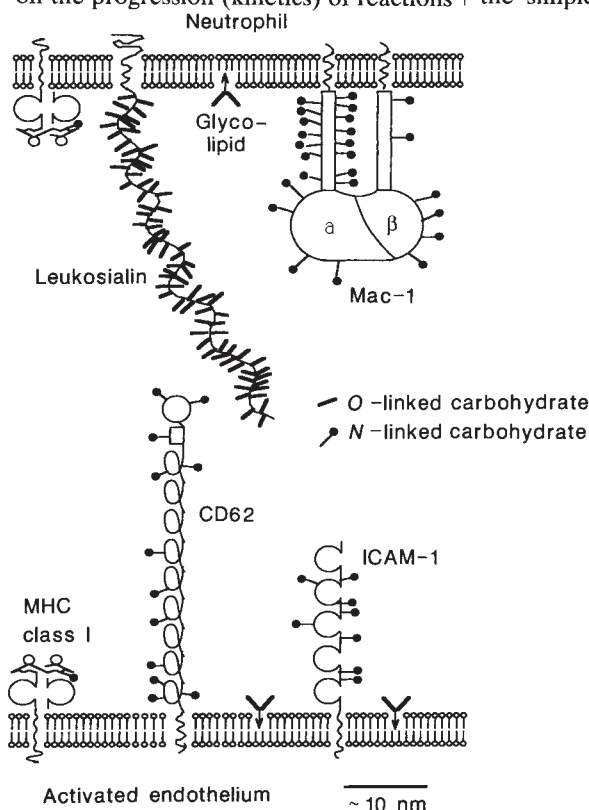
Alan F. Williams

THE concept of affinity dominates most thinking about complex biological reactions even though it is relevant only at equilibrium. An alternative is to focus on the progression (kinetics) of reactions

molecule must specifically recognize another and binding must persist for the time required for secondary events to occur. Most of us view these matters in the simplest of terms by considering a

bimolecular reaction $A+B \rightleftharpoons AB$. The concentration of the bound form at equilibrium is proportional to the association affinity constant, defined as $K_{eq} = [AB]/[A][B]$, and we tend to take this term as roughly indicating how long reactants will stay together. The higher the affinity constant the more chance of a biologically relevant reaction. This thinking is limited in scope, however, because the affinity constant applies only at equilibrium and gives no information about reaction rates. For the same affinity, the rates can readily differ by one thousand-fold. Take $K_{eq} = k_{on}/k_{off}$ where k_{on} and k_{off} are the intrinsic forward and backward rate constants. For antibodies reacting with chemical haptens, k_{on} values are often about $10^8 \text{ M}^{-1}\text{s}^{-1}$, whereas for antibody-protein interactions the figures are more commonly of the order of $10^5 \text{ M}^{-1}\text{s}^{-1}$ (ref. 3). With these values and a given K_{eq} of 10^9 M^{-1} the time for 50 per cent of a complex to dissociate ($t_{1/2} = 0.693/k_{off}$) will be 7 seconds for the hapten reaction, but 115 minutes for the antibody-protein reaction. These differences are of critical importance in understanding serological reactions³.

In the new papers the emphasis is on the forward reaction (on rate = k_{on} [A][B]). Foote and Milstein¹ have studied the maturation of the antibody response to the hapten oxazolone. Previously it was found that one group of antibodies predominated in the primary response, whereas in the secondary and tertiary responses a new set of antibodies emerged, along with clones from the original family. There were, however, no clear affinity differences in the two sets of antibodies. Why did a different family of antibodies emerge? In the current study, the kinetic parameters have been determined and it was found that many



Molecules of known or possible relevance to neutrophil-endothelial interactions. The molecules are shown roughly to scale at cell surfaces with dimensions as taken from refs 4, 8 and 9. Class I molecules of the major histocompatibility complex are shown for comparison of scale with the larger molecules. The integrin Mac-1 is shown with a structure based on that for the fibronectin receptor⁹ and LFA-1 can be expected to have a similar structure. Mac-1 and LFA-1 react with domains 1 and 3 respectively of ICAM-1 (ref. 10). The extracellular part of ICAM-1 is composed solely of immunoglobulin-related domains whereas that of CD62 contains, from the transmembrane sequence outwards, nine complement control protein modules followed by an epidermal growth factor repeat and the lectin-related domain¹¹. O-linked carbohydrate structures are shown only for leukosialin and the positioning of N-linked structures in Mac-1 is as in ref. 12.

and two new papers illustrate systems where the speed of a reaction may be of great importance. On page 530 of this issue¹, Foote and Milstein suggest that antibodies can be selected in a secondary antibody response because they show rapid binding kinetics; and, writing in *Cell*², Lawrence and Springer argue that rapid reaction rates determine which molecules mediate the initial interaction between neutrophils and endothelium during the early stages of inflammation.

Why is affinity such a popular parameter? To trigger complex events one

1. Uemura, Y. J. *et al.* *Phys. Rev. Lett.* **66**, 2665–2668 (1991).
2. Jorgensen, J. D., Hinks, D. G. & Fletcher, G. P. *Phys. Rev.* **B35**, 5365–5368 (1987).
3. Hebard, A. F. *et al.* *Nature* **350**, 600–601 (1991).
4. Uemura, Y. J. *et al.* *Nature* (in the press).
5. Oshima, K. *et al.* *Phys. Rev.* **B38**, 938–941 (1988).
6. Johnson, D. C., Tarascon, J. M. & Sienko, M. J. *Inorg. Chem.* **24**, 2598–2604 (1985).

of the new antibodies have k_{on} values that are ten times higher than for the secondary antibodies related to the first set of clones; that is, at a given concentration of antigen, the newly arising set of antibodies will react ten times faster than the others. Foote and Milstein argue that the antibodies are selected because of their faster reaction rates.

Lawrence and Springer² deal with the interaction of neutrophils with endothelium. During inflammation reactions *in vivo*, a cell is stopped in its flow through a capillary and rolls along the endothelial surface. The cell then adheres strongly to the endothelial surface before passing between cells into the tissue. Much work has been going on to identify the molecules that are involved in these interactions (for review and terminology see refs 4 and 5), and the importance of the integrins LFA-1 and Mac-1 binding to ICAM-1, a member of the immunoglobulin superfamily of adhesion receptors, was initially established. When ICAM-1 is inserted into a lipid bilayer on glass plates, activated neutrophils adhere well in static conditions and stay strongly attached if solution flow is initiated across the cells. But when the cells are applied under conditions of solution flow, neither resting nor activated neutrophils attach to ICAM-1 and it is clear that the integrins are not involved in the rolling reaction *in vivo*. Instead the initial reaction seems to be mediated by molecules of the selectin superfamily, and Lawrence and Springer establish a role for CD62 (also known as PADGEM or GMP-140). When CD62 is incorporated into lipid on the glass slide, a rolling reaction with unactivated neutrophils is initiated. Once this is established, it is believed that integrin binding

comes into play to trigger the later events provided the neutrophils also undergo activation with chemotactic agents.

Why is CD62 effective in catching neutrophils as they flow past whereas ICAM-1 is not? The ligand for CD62 is not precisely established but is thought to involve sialylated carbohydrate epitopes related to the Lewis x determinant (for review see ref. 5). Lawrence and Springer suggest that CD62 works because the rate of reaction with its carbohydrate epitope is faster than for interactions that involve protein epitopes. The k_{on} values for lectin-carbohydrate reactions have been measured at 5×10^4 and 5×10^5 M⁻¹ s⁻¹ for concanavalin A and *Ricinus communis* lectin, respectively⁶ and these are not higher than the constants for reactions between antibodies and proteins. However, Lawrence and Springer say that carbohydrate ligands may be found at the end of flexible chains leading to a more rapid rate in the complex situation of opposed cell surfaces. Another possibility is that the concentration of the carbohydrate epitopes is high at the cell surface and leads to the required rapid reaction.

On neutrophils the Lewis x carbohydrate can be expressed on glycolipids or on glycoproteins with O-linked or N-linked carbohydrate structures^{5,7}, and some of the molecules that may be involved are shown in the figure. An interesting molecule of possible relevance is leukosialin (CD43) which is expressed at high levels on neutrophils^{7,8}. This molecule extends to a length of 45 nm and carries about 70 O-linked carbohydrate structures on the 224 amino acids of its extracellular domain. It may be that CD62 reacts with

leukosialin or a similar structure at the periphery of the neutrophil surface and that the close clustering of carbohydrate epitopes yields a high local concentration leading to rapid on-rates in the reaction.

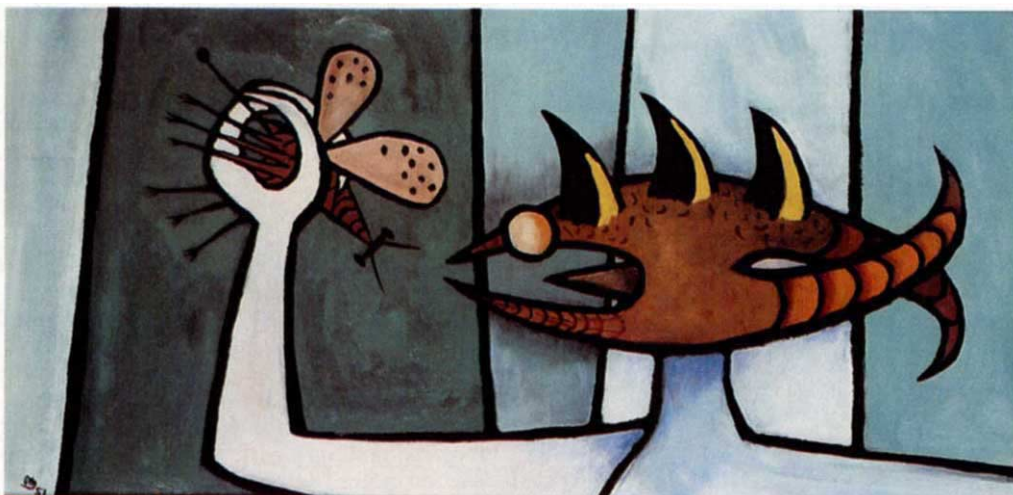
The concept of affinity has been invoked to explain all sorts of complex events; yet, as we better understand the real reactions, equilibrium considerations become largely irrelevant. Instead it is essential to understand the process of events in dynamic cell-cell interactions. Can molecules interact sufficiently quickly to be relevant? How many molecules must be engaged for a biologically relevant effect? How long must engagement occur for signals to be given? How important is the local organization of ligands at the cell surface? Which molecules act synergistically? It is questions such as these that will have to be answered if the complexity of cell interactions is to be resolved. □

Alan F. Williams is in the MRC Cellular Immunology Research Unit, Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK.

1. Foote, J. & Milstein, C. *Nature* **352**, 530-532 (1991).
2. Lawrence, M. B. & Springer, T. A. *Cell* **65**, 859-873 (1991).
3. Mason, D. W. & Williams, A. F. in *Handbook of Experimental Immunology*, Vol. 1, 4th edn (eds Weir, D. M. et al.) 38.1-38.17 (Blackwell Scientific, Oxford, 1986).
4. Springer, T. A. *Nature* **346**, 425-433 (1990).
5. Springer, T. A. & Lasky, L. A. *Nature* **349**, 196-197 (1991).
6. Podder, S. K., Suroli, A. & Bachhawat, B. K. *FEBS Letts* **85**, 313-316 (1978).
7. Fukuda, M., Carlsson, S. R., Klock, J. C. & Dell, A. *J. biol. Chem.* **261**, 12796-12806 (1986).
8. Cyster, J. G., Shotton, D. M. & Williams, A. F. *EMBO J.* **4**, 893-902 (1991).
9. Nermut, M. V. et al. *EMBO J.* **7**, 4093-4099 (1988).
10. Diamond, M. S. et al. *Cell* **65**, 961-971 (1991).
11. Johnston, G. I., Cook, R. G. & McEver, R. P. *Cell* **56**, 1033-1044 (1989).
12. Rosen, H. & Law, S. K. A. *Curr. Top. Microbiol. Immunol.* **153**, 99-122 (1989).

Reflections on a zoologist

DESMOND Morris is best known for his popularizations of behavioural studies, including *The Naked Ape* and *Manwatching* and even for an in-depth study of the psychology of football crowds, *The Soccer Tribe*. But with a catalogue of over 850 canvasses to his name, Morris is also a formidable artist. He admits a debt to the French surrealist Yves Tanguy, a mariner whose static images, influenced perhaps by desolate beachscapes, Morris contrasts with his own dynamic creations, dominated by "sexuality, violence and biological activity". He shared his first major exhibition in 1950 with his other mentor, Joan Miro. Shown here is one of a series of 'portraits' from 1951 (now in the collection of Dr John Godfrey), *The Zoologist*, in which stark lines and aggressive transformations intrude into Morris's comments on social functions. These paintings and many others appear in a retrospective text *The Surrealist World of Desmond Morris* by Michel Remy (Jonathan Cape, London, 1991; price £35).



A fondness for fungi

Robert M. May

ACCOMPANYING the rise in prominence of environmental issues of one kind or another has been the recognition that we know less than we thought we knew about the diversity of life around us. In particular, most readers will be aware that from recent studies of insects — by far the most diverse animal group by current counts — it seems that the number of species in the tropics may be far greater than previously thought¹⁻³. Exhibiting a degree of fungal chauvinism appropriate to his presidential address to the British Mycological Society, published in the June issue of *Mycological Research*⁴, David Hawksworth has now observed that fungi are the insects of the plant world. Hawksworth suspects there may be as many as 1.6 million species of fungi, rather than the 69,000 or so currently recorded.

Such an estimate depends of course on what is taken to constitute the fungi. Whittaker's five-kingdom system⁵, adopted soon after its appearance in 1969, gave them a kingdom to themselves, along with animals, plants, protists and monerans (roughly, bacteria). Advances in molecular systematics have resulted in further revisions to classificatory systems, and at present it is probably best to define fungi somewhat arbitrarily as "organisms studied by mycologists". This embraces phyla from the Kingdoms Protozoa (including Myxomycota), Chromista (including Oomycota), and Fungi in the strict sense (including those forming lichens). Hawksworth notes that some 64,000 species of fungi, thus defined, were recognized in 1983 (the date of publication of the most recent *Dictionary of the Fungi*⁶), and estimates some 5,000 have been added since, for a total of around 69,000 recorded up to 1990. He also surveys earlier estimates of the likely total, which range from around 100,000 to upwards of 300,000 (including one of 250,000 or more by Fries in 1825).

Hawksworth's new estimate is based on surveys of the ratio between species of vascular plants and species of fungi in well-studied areas. As for most groups, the most intensively documented region is the British Isles (a tradition that owes much to generations of clergy paying homage to their Creator in a useful way), where the vascular-plant:fungi ratio is around 1:6. Similar ratios are obtained for other regions — 1:4 for

Finland, 1:4 for Switzerland and 1:6 for alpine sedge communities. The ratio of 1:1 for the United States is discordant, but Hawksworth argues that these fungi are less well-studied than the Northern European ones (even more, ratios of 1:1 or 1:0.5 for India surely reflect lack of adequate description of the subcontinent's fungal species). Applying the 1:6

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

One in a million? Teliospores of *Diabole cubensis*, a new rust genus discovered in Mexico on the weed *Mimosa pigra*.

ratio to the global total of vascular plant species, currently estimated at 270,000, we arrive at 1.6 million species of fungi.

Hawksworth points out that this upward revision is in many ways conservative. First, it employs a modest estimate for the global total of vascular plants (estimates as high as 400,000 have been published). Second, it takes no explicit account of fungi that may be associated with insects; if the ratio of insect species to plant species is much higher in the tropics than in temperate regions, then projections based on multiplying temperate-zone plant:fungi ratios by total numbers of vascular plants will be too low. Third, even for the British Isles, vascular plant species are more completely recorded than are fungi — in each recent decade there has been roughly a 13 per cent increase in the number of recorded fungal species, and this trend shows no signs of slackening; on the other hand, the number of vascular plant species has remained roughly constant. Hawksworth estimates that his ratio will move from 1:6 to 1:7 in Britain by the turn of the century.

These estimates are valuable and helpful, but it seems to me that global estimates based on temperate-zone ratios of vascular plants to fungi may be too high. Turning back to insects, I believe Erwin's estimate¹ of a total of

around 30 million is almost surely excessive (my guess would be 5–8 million^{2,3,7}), essentially because one of the links in his chain of argument uses intuition based on temperate-zone data to estimate the fraction of insect species that are effectively specialized to a given tree species. This figure (20 per cent of herbivorous beetle species effectively specialized to the canopy of a given tree species) is likely to be an order of magnitude too high in the tropics, where qualitatively different patterns of tree species diversity are likely to compel

tropical insects to be more generalized with respect to host trees than are their temperate-zone cousins. Likewise, I think that fungal species may be more generalized with respect to hosts in the tropics, where greater diversity means essentially all host species are sparse (unlike many temperate-zone situations). If so, one cannot apply the vascular-plant:fungi ratio of 1:6 that pertains in Northern Europe to all the globe's vascular plants, most of which are found in the tropics. This is, of course, merely a conjecture, and one could alternatively argue that the appropriate ratio is even higher (rather than lower) in the tropics, owing to higher humidity,

more complex architecture of vegetation, or other factors.

Be that as it may, other facts surveyed by Hawksworth point towards his estimate being on the high side. If there really are 1.6 million species of fungi, of which only 69,000 have been recorded, then 96 per cent of the fungi found in any newly studied area may be expected to be previously unknown, especially in tropical regions. Hawksworth summarizes a variety of recent studies from Brazil, Sierra Leone, East Africa, the Lesser Antilles, Papua New Guinea, Malaysia, India, as well as more northerly and southerly regions. The proportion of previously undescribed species in particular groups in these studies is typically around 15–30 per cent, with a few higher numbers (30 new species among 35 *Oropogon* species from new world sites⁸; 100 species among 140 bolete species from Malaysia⁹). Taken by themselves, these numbers suggest the global total of fungal species is less than one million, and more like a few hundred thousand. But there could easily be a tendency to focus such studies on the better-known groups of fungi, even in the tropics, for understandable reasons, which would mean they cannot be used as an independent check on Hawksworth's estimate.

Hawksworth also emphasizes the problems of codifying information about

Illuminated view

WEAK binding processes are hard to analyse because the free ligand is performed always in large excess over the bound. C. L. Poglitsch *et al.* (*Biochemistry* **30**, 6662–6671, 1991) now demonstrate the virtues of a new method which they use to quantitate the weak interaction between an IgG and an Fc receptor *in vitro*. The receptor is incorporated into liposomes, which are then deposited on a quartz plate and fused into a continuous bilayer. This is placed in contact with a solution of fluorescein-labelled antibody and examined by fluorescence, excited by a laser at a glancing angle to give total internal reflection. The light penetrates less than 100 nm into the solution and so excites only surface-bound fluorophores. Binding isotherms of enviable precision deliver an association constant of a mere $3 \times 10^5 \text{ M}^{-1}$.

Unnatural gas

Cool astrophysical clouds are known to contain all kinds of organic molecules, so it comes as a bit of a surprise to see J. H. Lacy *et al.* (*Astrophys. J.* **376**, 556–560, 1991) announcing the first detection of methane, the simplest saturated hydrocarbon, outside the Solar System. Many other simple hydrocarbons have been found, but the weakness of methane's molecular lines allowed it to elude detection until now. Lacy *et al.* distinguish both gaseous and solid methane (the latter presumably a mantle on dust grains) in the spectra of molecular clouds around young stars. The presence of either form is not hard to explain, but their coexistence is more problematic. There is less gas than would be expected if the grain mantles evaporate by stellar irradiation. The authors speculate that gaseous methane must be vulnerable to gas-phase reactions in the dense molecular clouds.

Rats' tales

RATS resistant to warfarin pay a huge price for the privilege. Warfarin is an anticoagulant, thought to work by inhibiting an enzyme called vitamin K epoxide reductase which is involved in the production of blood clotting factors. The drawback is that mutant enzymes are almost as unresponsive to vitamin K, their co-factor, as to warfarin itself. The result is that the protected rats are deficient in vitamin K. They are also smaller than their susceptible cousins, but expenditure of extra energy in patching up the defect may not be the reason. P. Smith *et al.* (*Functional Ecology* **5**, 441–447, 1991) note that a reduced version of vitamin K is essential for maintaining high concentrations of the bone growth promoter osteocalcin, and suggest that it is this aspect of vitamin K deficiency that stunts growth.

the known fungal species, much less those yet unknown. The International Mycological Institute at Kew, for instance, holds a reference collection of about 31,500 species, which occupy 1,456 metres of shelving in compactors. An equivalent collection of Hawksworth's global total of 1.6 million species would run to over 70 km of shelving. The need to computerize the data, preferably on CD-ROM disks with graphic and picture-storing capability, is evident, even though the costs are high in comparison with the traditions of this discipline¹⁰ (but not in comparison with other areas seen as 'big science'). As for conservation of species endangered by habitat loss or other environmental changes, it is sufficient to note that *ex situ* collections of living fungal cultures include fewer than 12,000 species, or around one sixth of known species.

Fungi lack the appeal of the panda, or even of the average beetle, but they are key players in most ecosystems, especially as decomposers and soil-developers.

PLATE TECTONICS

About turn for supercontinents

Chris J. H. Hartnady

THE supercontinent Pangaea, which existed about 200 million years ago, was formed by the (not much) earlier collision of two smaller land masses, Gondwanaland to the south and Laurasia to the north. Was there also an ancient supercontinental ancestor to these blocks; and if so, what did it (or they) actually look like? These are among the issues addressed in a series of related papers by Moores¹, Dalziel², and Hoffman³, with fresh perspectives on the evolution of Gondwanaland and the configuration of its precursor supercontinent during the Late Proterozoic era (900–530 million years ago). The answers that are emerging may surprise — not least the suggestion that Gondwanaland arose when its predecessor "turned inside-out"³ on the spherical surface of our planet. This suggests that continental rifting and accretion by continent-continent collision are two facets of the same plate-tectonic process.

The essential palaeogeography of Gondwanaland was established in 1937 by Du Toit⁴, and a geological wall-map now exists of the continent⁵. A common view is that it originated by rifting of a unitary Proterozoic supercontinent⁶, which also spawned its counterpart in the Northern Hemisphere, Laurasia. But the idea that, through most of geological time, a single primal supercontinent drifted more or less rigidly over the surface of the globe is open to serious challenge. Instead, the continuous op-

Hawksworth discusses their significance in the evolution of organic diversity, suggesting that they were crucial to land colonization by plants, and — especially through the development of mutualisms — to the subsequent radiation of other groups, particularly vascular plants and many insects. Like Rodney Dangerfield, they deserve more respect. □

Robert M. May is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

1. Erwin, T. L. *Coleopt. Bull.* **36**, 74–82 (1982).
2. Stork, N. E. *Biol. J. Linn. Soc.* **35**, 321–337 (1988).
3. May, R. M. *Proc. R. Soc. Lond.* **B330**, 293–304 (1990).
4. Hawksworth, D. L. *Mycol. Res.* **95**, 641–655 (1991).
5. Whittaker, R. H. *Science* **163**, 157–160 (1969).
6. Hawksworth, D. L. *et al.* *Ainsworth and Bisby's Dictionary of the Fungi*, 7th edn (Commonwealth Mycological Institute, Kew, 1983).
7. Hodgkinson, I. D. & Casson, D. *Biol. J. Linn. Soc.* **43**, 109–109 (1991).
8. Esslinger, T. L. *Systematic Botany Monographs* **38**, 1–111 (1989).
9. Corner, E. J. H. *Boletus in Malaysia* (Government Printing Office, Singapore, 1972).
10. Hawksworth, D. L. & Mound, L. A. in *The Importance of Microorganisms and Invertebrates in the Maintenance of Sustainable Agriculture* (ed. D. L. Hawksworth) 230–250 (CAB International, Wallingford, 1991).

eration of plate tectonics over the greater part of Earth history has probably led, via the so-called 'Wilson Cycle', to the episodic disruption and re-assembly of continental fragments into a number of varied mosaical configurations at different times⁷. Indeed, the alternative view of Gondwanaland's origin is that it formed not by rifting from a larger whole, but by a collision between the putative, smaller West Gondwanaland and East Gondwanaland blocks⁸. The high-grade metamorphic Mozambique Belt of East Africa and Madagascar is usually considered to be the zone of collisional mountain construction (orogeny) within which the cryptic suture between the two halves of Gondwanaland is concealed.

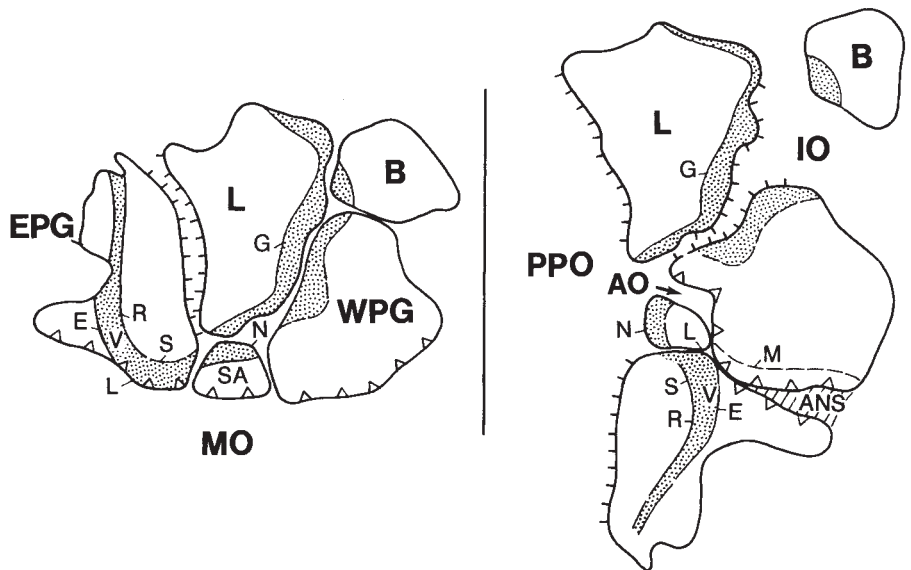
The hypothesis of a geological connection between southwest United States and East Antarctic ("SWEAT")¹ was inspired by an earlier matching of late Precambrian rocks between western Canada and South Australia on stratigraphic, metallogenetic and palaeomagnetic grounds⁹. Before their separation by sea-floor spreading, only about 50 million years (Myr) ago, continental East Antarctica was previously attached to the stable cratonic shield of Australia. Therefore, when the Pacific-facing margins of a combined East Antarctica–Australia block are reconstructed in their possible pre-Gondwana configuration against the conjugate western margin of the North America continental

shield or (Laurentia, not to be confused with Laurasia, of which it is only a part), there should be a coherent pattern of ancient Archaean nuclei and Early- to Mid-Proterozoic orogenic belts that match from one block to the other¹.

Moore's notes¹ that his hypothesis "implies that Gondwana and Laurentia rifted away from each other on one margin and collided some 300 Myr later on their opposite margins to form the Appalachians". This statement relates the origin of the modern Pacific Ocean to the demise of the earlier Iapetus ocean, which finally closed completely along a suture somewhere between the Appalachian chain of North America and the Mauretanide chain of North Africa. Dalziel² widens the perspective on this, by postulating the closure of an older 'Mozambique' Ocean as the counterpart to the opening of the juvenile Iapetus ocean. In this process, one part of Gondwana pivots through 180° or more about the other part, while both remain in contact as if being folded around a hinge-zone. This stage is followed by the opening of the (proto-) Pacific ocean and the concurrent closure of Iapetus, as Laurentia pivots and partially circles around a united Gondwanaland.

Hoffman's contribution³ — which started as a referee's comment on the papers by Moore and Dalziel — poses the title question: "Did the breakout of Laurentia turn Gondwanaland inside-out?" But is the precursor supercontinent correctly named? Its postulated palaeogeography (see figure) was certainly very different, and includes components, like Laurentia, which — with some reservation^{7,10} — have not previously been included in the Gondwana realm. Shouldn't it rather be named distinctively? My preference is for Ur-Gondwanaland, which indicates that the supercontinent was a primitive, original or earliest manifestation of Gondwana. The alternative term proto-Gondwanaland, which has a similar meaning, could then be reserved for the two halves that subsequently collided to form Gondwanaland proper. (There may also be waggish suggestions that the pre-Gondwana supercontinent should be called Laurentiocentria or Hoffmanian.)

Before its "breakout", Laurentia occupied a roughly central, keystone position within Ur-Gondwanaland (see figure): to its east (in a present-day American geographical reference frame) lay a region that was later to rift away as West proto-Gondwanaland (WPG); to its west lay the region that was to become East proto-Gondwanaland (EPG). Although epitomizing the essential kinematics of the continental movements, Hoffman's title (although not his text) seems to imply something about the timing of the



Left (modified from Hoffman³), Ur-Gondwanaland in its pre-750 Myr-old configuration with Laurentia (L) in a keystone position, flanked on its 'west' by East proto-Gondwanaland (EPG), on its 'east' West proto-Gondwanaland (WPG) and Baltica (B). The southern African (SA) block is situated in a hinge position between EPG and WPG, 'south' of L. The Grenvillian belts (stippled) probably formed during Ur-Gondwanaland assembly by collision between the SA-WPG-B block and the larger part of the EPG-L block. Locations of the Grenville (G), Namaqua (N), Lurio (L), Sverdrupfjella (S), Vijayan (V), Eastern Ghats (E) and Rayner (R) segments are marked. Hatched lines mark the internal rift boundaries of the incipient proto-Pacific Ocean between L and EPG. Barbed lines mark the external boundary between Ur-Gondwanaland and the Mozambique Ocean² (MO). Right, the 650–600 Myr-old configuration Ur-Gondwanaland has been turned inside out (extraverted) around the southern African hinge-block. The Grenvillian belts (G, N, L, S, V, E and R; stippled) have been considerably straightened from their original arcuate form. The former Mozambique Ocean has been closed along the Mozambique metamorphic belt (M) and new crust has been added by island-arc accretion in the wedge-shaped Arabian-Nubian Shield (ANS). With continued convergence across the Mozambique Belt between WPG and EPG the Adamastor Ocean (AO) opens as an arm of the new proto-Pacific Ocean (PPO). The Iapetus Ocean (IO) begins to open between Laurentia (L), Baltica (B) and Gondwanaland.

process, namely, that the breakout and the turning inside-out were concurrent. The start of the transformation is estimated at 700 Myr ago, and Gondwanaland is shown in its modern configuration 500 Myr ago. Dalziel's model², however, envisages a temporal succession of distinct stages during the transition from Ur-Gondwanaland; first the folding of the EPG and WPG fragments, with Laurentia still rigidly attached to the 'eastern' part; then the separation of Laurentia from an already united Gondwanaland; finally their re-convergence in Pangaea. But the occurrence of all this is constrained to the 300-Myr interval from the start of the Cambrian Period, 570 Myr ago, to 250 Myr ago. Is this really enough time to close the late Proterozoic ocean along the Mozambique Belt while opening up Iapetus, and then close Iapetus while opening up the proto-Pacific Ocean? A quantitative estimate of the minimum rates of spreading/convergence required for these tectonic transformations would be useful (I.W.D. Dalziel, personal communication).

In a speculation⁷ about Laurentia-Gondwanaland relationships — prompted by Moore's remarks during a 1984 field trip to the Gariep Belt along the border of Namibia and South Africa about an

uncanny resemblance of the Late Proterozoic stratigraphic sequence there with the Pahrump Group in Death Valley, California — I envisaged the rifting of EPG from western Laurentia about 800 Myr ago, well before its later collision with WPG, which was then (erroneously, I now think) referred to as the early Palaeozoic. In this model, the closure of the Mozambique Ocean is the counterpart to the opening of a proto-Pacific ocean, and not of Iapetus. The Southern African block probably participated in the motion of EPG, being separated from WPG by a proto-Pacific arm called the Adamastor Ocean¹¹, which may have had some oceanic connection with the shrinking Mozambique basin. The Mozambique Belt granulite metamorphism, about 715 Myr old¹², may therefore be related to the main collision event between EPG and WPG, which was the culmination of the turning inside-out process. Rifting between the western (Andean) margin of WPG and the eastern (Appalachian) margin of Laurentia, and hence the opening of Iapetus, commenced only 600 Myr ago¹³ after Gondwanaland had already been united by the main EPG-WPG collision.

The Mozambiquian granulite age could, however, be unrelated to EPG-

WPG collision, which might only have occurred about 100 Myr later. It might date a magmatic underplating episode beneath an Andean-type continental volcanic arc, or even an arc-continent collision at the exterior margin of Ur-Gondwanaland. Nevertheless, the evidence seems to indicate that the Ur-Gondwanaland transformation started much earlier, probably⁷ when EPG pivoted away from western Laurentia 800 Myr ago, and that Laurentia remained rigidly connected to WPG for a further 200 Myr. The best available palaeomagnetic data show that Gondwanaland had already turned inside-out by 570 Myr (A. G. Smith, personal communication), and that Laurentia was separated from it by the Iapetus ocean. There are, however, some models (for example, ref. 14) that show that EPG and WPG were still separated at the time of the Precambrian-Cambrian boundary, which may have been as recent as 520–540 Myr ago.

All three new papers imply that the well-known Grenville Province of eastern and southern North America continues directly into rocks of similar Mid-Proterozoic age (1,000 Myr old) in Dronning Maud Land of East Antarctica. Belts of Grenvillian age occur widely on all Gondwanaland fragments, and share many similar features. They are characterized particularly by extensive surface exposures of high-grade, granulite-facies metamorphic rocks, and large, massif-type intrusions of anorthosite and charnockite. These distinctive lower-crustal metamorphic and igneous associations are usually explained by deep exhumation of crust that was overthickened during earlier continent-continent collision. The extensive pattern of Grenvillian orogenic belts, particularly its boundary with older terranes, provides a convenient system of piercing points — where the boundary is trans-

ected and displaced by edges of younger belts — for tectonic reconstructions of the Ur-Gondwanaland jigsaw puzzle.

The new hypotheses about the changing shapes and associations of supercontinents have a wider crustal evolutionary implication for the Wilson Cycle. It seems that, on the spherical Earth, plate motions established during the early separation of two continental masses are rarely, if ever, exactly reversed so that the former rifted margins are rejoined by collision. Instead disrupted supercontinents are re-assembled in a radically different configuration, after two or more fragments have broadly looped around each other, or have been 'extra-verted' — folded back so as to turn inside-out. But pre-Mesozoic plate tectonics is not merely about elaborating the history of palaeocontinental drift, using orogenic jigsaw patterns and

palaeomagnetism. It is also about the reconstruction or resurrection of long-lost ocean basins, like Iapetus and Adamastor. Earth scientists should pay more attention to the underlying geodynamic significance of what might be called supercontinental extraversion. The main driving force in plate tectonics is, after all, subduction of oceanic lithosphere. One of the least secure and probably most misunderstood aspects of the Wilson Cycle is subduction initiation¹⁵, or how opening Atlantic-type basins become converted to shrinking Pacific-type basins. The remarkable gyrations of Gondwanaland and its surrounding oceans may contain important keys to the unlocking of this question. □

Chris J. H. Hartnady is in the Precambrian Research Unit, Department of Geology, University of Cape Town, Rondebosch 7700, South Africa.

STRUCTURAL BIOLOGY

Complex behaviour

Daniela Rhodes and John W. R. Schwabe

ON page 497 of this issue, Luisi *et al.*¹ report the crystal structure of a complex between DNA and a hormone receptor that acts as a transcription factor. Such high-resolution crystal structures reveal the chemistry of the protein-DNA interface, which in turn can shed light on the structural basis of the sequence-specific recognition that is central to the control of gene transcription.

Much of our knowledge of sequence-specific recognition came originally from the helix-turn-helix motifs found in prokaryotic transcriptional regulators². But recently the crystal structures of two eukaryotic protein-DNA complexes have been determined (a homeodomain from *Drosophila*³ and a zinc-finger motif from mouse⁴). Luisi *et al.* now present the third crystal structure of a eukaryotic protein-DNA complex, representing yet another class of DNA-binding motif. This is the complex of the DNA-binding domain (DBD) of the glucocorticoid receptor with DNA. This transcription factor is a member of a superfamily of hormone-activated nuclear receptors which play a central role in the control of eukaryotic gene expression. A striking feature of this family is that their DBDs are conserved and also recognize related DNA sequences, thus presenting a particularly intriguing problem: how is specific recognition achieved?

The receptor DBD consists of about 80 amino acids and contains two sets of four cysteine residues. Each set coordinates a zinc ion with tetrahedral geometry, and thus at first sight the domain appeared to resemble the classic 30-residue zinc-finger motif first seen in

TFIIIA⁵. But the three-dimensional structure of the two motifs, determined using two-dimensional NMR spectroscopy, showed that the receptor DBD is structurally distinct from the zinc-finger motif^{6,7}. Consecutive TFIIIA-type zinc-fingers form independent mini-domains, whereas the DBD of hormone receptors forms a single domain which has a globular fold and which is similar to the helix-turn-helix motif in that an exposed DNA-recognition helix is buttressed through interaction with a second helix. The structures of the DBD of both the glucocorticoid⁶ and oestrogen⁷ receptors have been determined. These structures are very similar and thus establish a common architecture for DNA recognition. The DNA-recognition helix and a second amphipathic α helix are folded together, so that they cross at right angles at their midpoints to give the protein an extensive hydrophobic core. A zinc ion holds the peptide loops at the N terminus of each helix.

Luisi *et al.* have determined the structure of the DBD of the glucocorticoid receptor bound to two different DNA-binding sites, each of 18 base pairs. The low-resolution native complex consists of a consensus, functional palindromic glucocorticoid-response element (GRE) of two hexameric half-sites separated by three base pairs (GRE_{S3}) on which is bound a DBD dimer. The resolution of this structure is insufficient to observe the side-chain conformation. The higher-resolution (2.9 Å) complex has a perfectly symmetrical GRE with a non-native spacing between the two half-sites of four base pairs (GRE_{S4}). The DBD

1. Moores, E. M. *Geology* **19**, 325–428 (1991).
2. Dalziel, I. W. D. *Geology* **19**, 598–601 (1991).
3. Hoffman, P. F. *Science* **252**, 1409–1412 (1991).
4. Du Toit, A. L. *Our Wandering Continents* (Oliver & Boyd, Edinburgh, 1937).
5. De Wit, M., Jeffery, M., Bergh, H. & Nicolaysen, L. O. *Geological Map of Sector of Gondwana Reconstructed to their Disposition ~ 150 Ma* (American Association of Petroleum Geologists, Tulsa, 1988).
6. Piper, J. D. A. *Palaeomagnetism and Continental Evolution* (Open University Press, Milton Keynes, 1987).
7. Hartnady, C. J. H. *S. Afr. J. Sci.* **82**, 251–254 (1986).
8. McWilliams, M. O. in *Precambrian Plate Tectonics* (ed. Kröner, A.) 649–687 (Elsevier, Amsterdam, 1981).
9. Bell, R. T. & Jefferson, C. W. in *Proc. Pacific Rim Congress '87* 39–50 (Australian Institute of Mining and Metallurgy, Parkville, Victoria, 1987).
10. Bond, G. C., Nickeson, P. A. & Kominz, M. A. *Earth planet. Sci. Lett.* **70**, 325–345 (1984).
11. Hartnady, C., Joubert, P. & Stowe, C. *Episodes* **8**, 236–244 (1985).
12. Maboko, M. A. H., Boelrijk, N. A. I., Priem, H. N. A. & Verdurmen, E. A. *Precamb. Res.* **30**, 361–378 (1985).
13. Williams, H. & Hiscott, R. N. *Geology* **15**, 1044–1047 (1987).
14. Kirschvink, J. L., Margaritis, M., Ripperdan, R. L., Zhuravlev, A. Yu. & Rozanov, A. Yu. *GSA Today* **1**, 69–71, 87–91 (1991).
15. Mueller, S. & Phillips, R. J. *J. geophys. Res.* **96**, 651–665 (1991).

again binds as a dimer, but the consequence of the 'incorrect' spacing is that the two monomers in the dimer do not bind equivalently. Luisi *et al.* propose that one monomer binds specifically, whereas the other binds non-specifically.

The DBD, which is monomeric in solution^{6,7}, dimerizes when it binds to DNA, placing the recognition helices of each DBD in adjacent major grooves of the DNA. The two DBD domains make several protein-protein contacts where they meet between the two half-sites. These contacts appear to be the same in the two crystal structures. This region of protein is flexible in solution^{6,7} and becomes structured only on binding to DNA, when it makes contacts with both the DNA and the other DBD. This accounts for the cooperative binding of two DBD domains to a GRE with the native three-base-pair spacing⁸.

This new work confirms the arrangement of the protein-DNA complex that was predicted from NMR structures together with mutagenesis and biochemical studies^{6,7}. In addition, this work reveals the precise chemistry of the protein-protein and the protein-DNA interfaces. Owing to the low resolution of the native GRE_{S3}-DBD complex, we must rely on the interactions seen on the 'specific' side of the high-resolution structure (DBD-GRE_{S4}). The protein is anchored to the phosphate backbone making a total of seven contacts on either side of the major groove. The DNA-recognition helix protrudes into the major groove, making three contacts to the base pairs. The DNA has a B-like structure, but at the 'specific' site the major groove is widened by about 2 Å to accommodate the recognition helix. The fitting of a DNA-recognition helix into the major groove of the DNA double helix is a recurring feature of sequence-specific recognition. (On the 'non-specific' side of the GRE_{S4} complex, the contacts made are rearranged and fewer in number.)

Many of the amino acids involved in DNA recognition, formation of the dimer interface, and maintenance of the three-dimensional structure have been identified by mutagenesis. In particular, three amino acids are responsible for discriminating between the glucocorticoid- and oestrogen-receptor half-sites (which differ by only two base pairs)⁹⁻¹¹. These amino acids are glycine, serine and valine in the glucocorticoid receptor. Remarkably, only one of these residues makes contact with the DNA in the glucocorticoid receptor DNA complex. This is a van der Waals contact between a valine (alanine in the oestrogen receptor) and the methyl group of a thymine. Although the glycine is unlikely to make contact with the DNA, it seems surprising that the serine residue,

which can make hydrogen bonds, is not involved.

An intriguing feature of the structure of the DBD-GRE_{S4} complex is that the protein-protein contacts at the dimer interface can override the potential specific protein-DNA contacts at the second site in the GRE, yet they are not strong enough to hold the protein as a dimer in solution^{6,7}. There are two possible explanations for this. First, the protein-protein dimer interface might be induced and stabilized by interactions with the DNA backbone. Second, the high concentrations of protein and DNA in the crystallization conditions (that is, above the concentration of nonspecific binding) and crystal-packing forces might favour dimer formation.

The specific-nonspecific character of this complex arises because of the presence of an extra base pair between the two half-sites in the GRE. To obtain specific binding at both half-sites, the native spacing of three base pairs is required. Thus, as Luisi *et al.* note, the information contained in the spacing of the two half-sites is an integral part of the binding-site 'code', which is 'recognized' through the formation of a protein-dimer interface. This is especially important for other members of the nuclear-receptor family (the oestrogen, thyroid hormone, retinoic acid and vitamin D receptors), which recognize the same half-site sequence, but with different orientations and/or spacings^{12,13}. For these receptors, orientation and spacing are the only means of discrimination.

So the new crystal structure gives us a detailed understanding of how the glucocorticoid receptor achieves specific binding to its target DNA binding site. But we must await the determination of similar structures for other members of the nuclear-receptor family, particularly those that bind to directly repeated half-sites, before we have the complete story. □

Daniela Rhodes and John W. R. Schwabe are at the Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK.

1. Luisi, B. F. *et al.* *Nature* **352**, 497-505 (1991).
2. Harrison, S. C. & Aggarwal, A. K. *Rev. Biochem.* **59**, 933-969 (1990).
3. Kissing, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. *Cell* **63**, 579-590 (1990).
4. Pavlath, N. P. & Pabo, C. O. *Science* **252**, 809-817 (1991).
5. Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **4**, 1609-1614 (1985).
6. Härd, T. *et al.* *Science* **249**, 157-160 (1990).
7. Schwabe, J. W. R., Neuhaus, D. & Rhodes, D. *Nature* **348**, 157-160 (1990).
8. Tsai, S. Y. *et al.* *Cell* **55**, 361-369 (1988).
9. Madar, S., Kumar, V., De Verneuil, H. & Chambon, P. *Nature* **338**, 271-274 (1989).
10. Danielson, M., Hinck, L. & Ringold, G. M. *Cell* **57**, 1131-1146 (1989).
11. Umesono, K. & Evans, R. M. *Cell* **57**, 1139-1146 (1989).
12. Umesono, K. *et al.* *Cell* **65**, 1255-1266 (1991).
13. Näär, A. M. *et al.* *Cell* **65**, 1267-1279 (1991).

The highway code

THE supermarket bar-code reader is diversifying into all sorts of routine tasks of sorting and identification. Daedalus now plans to import it into road transport. He is devising a downward-looking bar-code reader that scans the road under a vehicle, and a range of stencils for laying bar-code messages on the roads. The idea is to automate and extend the conventional system of road signs.

Road navigation would be transformed. Computerized routing programs already exist, but are useless if you don't know at every instant where you are. Starry-eyed technomaniacs have suggested fitting every car with its own satellite navigator; but a simple bar-code road marking at each road junction would solve the problem far more directly. The vehicle's computer would not have to worry about absolute positions, and could find its way around the known road-network with absolute assurance. The most complicated and badly sign-posted cities, like London or Tokyo, would lose their navigational terrors. Bus and taxi drivers would no longer need elaborate training, and travelling salesmen could confine their mathematical efforts to routing theory.

Even finer details could easily be encoded. Lane-identifiers could warn the driver to get into the correct lane for his destination, instead of having to veer across many lanes of enraged traffic for lack of local knowledge. Where appropriate, the bar-codes could even include the house numbers along the road. The familiar British ordeal of trying to drive safely, especially at night, while scanning the side of the road for distant, tiny, unpredictably placed, and quite possibly non-existent house numbers, would be mercifully ended.

Bar-coded roads would be safer, too. The local speed limit could be encoded, with the car's computer programmed to display increasingly disapproving messages if it was exceeded. Temporary bar-codes could be laid down to warn of diversions and lane closures. Even at night or in thick fog, an emergency bar-code mat laid on the road would reliably warn of an accident or emergency ahead. Bar-coded 'traffic lights', perhaps an LED or liquid-crystal display embedded in the road, could give far more information than three coloured lamps; if the driver failed to react to them, the vehicle's computer could set off a sequence of dire warnings. But the direst warning of them all could be set off by any bar-code at all — by being read backwards. This proof that the car was on the wrong side of the road should perhaps slam the brakes into an emergency crash-stop.

DAVID JONES

Fullerenes from coal

SIR — Given their many potential applications¹⁻⁴, it is likely that C₆₀ and C₇₀ fullerenes⁵ will be required in macroscopic quantities. So far such quantities have been produced from the heating or arcing of graphite^{6,7}. But graphite is already a value-added carbon, costing \$1,000–5,000 per tonne. Previously we have detected C₆₀ and higher fullerenes during laser ablation of a brown coal in an ion cyclotron mass spectrometer⁸⁻¹⁰. We have now prepared macroscopic quantities of fullerenes from coal-derived coke, a relatively cheap and abundant source.

A coking coal from the Goonyella seam, Australia, was demineralized according to published procedures¹¹. Three fractions of this coal with mineral matter contents of 0.7, 2.9 and 7.9 per cent, respectively were obtained. Each coal fraction was finely ground and heated in argon for 24 hours at 395 °C to form a rod of approximately 18 mm (diameter) × 45 mm (length). The rod was converted to coke by further heating at 1,200 °C in argon for 5 h. The final coke rod has an electrical resistance of between 1.5 and 3 ohms along its length. In a typical experiment, two coke rods were subjected to electrical arcing at 24 V and an a.c. current of 105–110 A, in a manner similar to that of Haufier *et al.*⁶, under a 250-torr helium atmosphere. The soot collected was Soxhlet-extracted with toluene. The yield data of the toluene-soluble product and the duration of arcing for the different coal fractions are listed in the table. Infrared spectroscopy of the toluene-extracted product reveals absorption peaks characteristic of C₆₀ and C₇₀ fullerenes in ratios of about 10:1, similar to those reported with graphite as the source⁷. These ratios have been confirmed by ¹³C NMR. The C₆₀ and C₇₀ fullerenes were separated by chromatographic techniques¹².

The presence of mineral matter in the coal does not inhibit fullerene formation

from coke, although the amount of mineral matter does seem to have an effect on the total yield. At a level of 2.9 per cent mineral matter, an optimal yield of 8.6 per cent was obtained, comparing favourably with the yield of 9.3 per cent obtained from graphite under identical conditions. Fullerenes were also produced from coke derived from Coalcliff middlings coal (3 per cent yield with a d.c. current and 2 per cent with a.c.) and Newvale vitrinite (6 per cent yield d.c.). Thus it seems likely that other conductive coke materials will be able to act as a source of fullerenes. The use of a d.c. current rather than a.c. seems to increase markedly the yield of fullerenes

from both graphite and coal. This is probably because of the more stable arc and the relatively higher mean power when d.c. is used. X-ray diffraction of the coke rods shows very little evidence of graphitization: Coking occurs readily at 1,200 °C, but graphitization does not occur fully until 2,500 °C (ref. 13). With the present cost of coke at about \$500 per tonne, its use as an industrial source of fullerenes would greatly improve the economics of production.

LOUIS S. K. PANG

ANTHONY M. VASSALLO

MICHAEL A. WILSON

CSIRO Division of Coal and
Energy Technology,
PO Box 136, North Ryde,
New South Wales 2113,
Australia

EXPERIMENTAL CONDITIONS AND YIELDS OF CRUDE FULLERENES FROM COKE AND GRAPHITE

Source carbon	Coal mineral matter content (wt %)	Current type	Duration of arcing (h)	Soot yield (g) (gh ⁻¹)	Toluene solubles (g)	Percentage yield from soot (wt %)
Coke (Goonyella)	7.9*	a.c.	4	3.20 0.80	0.18	5.6
Coke (Goonyella)	2.9†	a.c.	4	1.74 0.435	0.15	8.6
Coke (Goonyella)	0.7†	a.c.	1	1.09 1.09	0.025	2.3
Graphite	NA	a.c.	4	1.07 0.28	0.100	9.3
Graphite	NA	d.c.	2	1.91 0.96	0.31	16.2

* From as-received coal, air dried and contains 1.3 per cent moisture.

† From demineralized coal.

NA not applicable.

Penguins shed stomach linings

SIR — As a member of the 1990–91 Dutch Antarctic expedition, I observed chinstrap penguins (*Pygoscelis antarctica*) vomiting their stomach linings. On closer inspection, the colony was strewn with shed linings, and a penguin regurgitating its stomach wall on its way to the sea to feed was a common sight. This habit, which has never been reported, is not seen in other *Pygoscelis* species, despite almost identical food and feeding habits.

I could think of two reasons for my observation. First, the stomach wall might need adjustment when birds have to feed chicks. The phenomenon would then be seen only near hatching time, and might therefore have remained undetected previously. By dye-marking birds, I established that regurgitation happens in birds with chicks of all ages, but also in non-breeders, throughout the chick-rearing season.

Second, birds could accumulate fluoride in their stomach walls, avoiding excessive fluoride uptake by regular

shedding. Antarctic krill contains toxic levels of fluoride, concentrated in the exoskeleton to up to 3,000 p.p.m. dry weight (dwt). Many antarctic animals feed on krill, and various authors²⁻⁴ have reported on fluoride levels and fluoride turnover in antarctic animals. Penguins and seals accumulate fluoride in their bones up to 10,000 p.p.m. dwt, and keep levels in their soft tissues below 10 p.p.m. dwt. Some chinstrap stomach walls were analysed by Boris Culik, and appear to contain up to 330 p.p.m. dwt, the highest fluoride levels ever found in soft tissue. In *P. adeline* stomach walls, which are not regurgitated, Culik found only slightly raised levels, up to about 20 p.p.m. dwt, which he attributed to contamination with krill fragments. Despite this difference, daily shedding of stomach walls with concentrations around 300 p.p.m. dwt would remove only an insignificant part (about 0.1 per cent) of the daily amount of fluoride

1. Adelung, D., Buchholz, F., Culik, B. & Keck, A. *Polar Biol.* **7**, 43–50 (1987).

2. Schneppenheim, R. *Meeresforschung* **28**, 179–182 (1980).

3. Culik, B. *Comp. Biochem. Physiol.* **88A**, 229–233 (1987).

4. Culik, B. *Polar Biol.* **7**, 179–187 (1987).

- Stoddart, J. F. *Angew. Chem. Int. Ed. Engl.* **30**, 70–71 (1991).
- Haddon, R. C. *et al.* *Nature* **350**, 320–322 (1991).
- Hebard, A. F. *et al.* *Nature* **351**, 600–601 (1991).
- Burch, W. M., Sullivan, P. J. & McLaren, C. J. *Nucl. Medical Commun.* **7**, 865–871 (1986).
- Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **338**, 162–163 (1985).
- Haufier, R. E. *et al.* *J. Phys. Chem.* **94**, 8630–8636 (1990).
- Kratschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347**, 354–358 (1990).
- Greenwood, P. F. *et al.* *Am. Chem. Soc. Div. Fuel Chem. Prepr.* **34**, 773–779 (1989).
- Greenwood, P. F. *et al.* *Fuel* **69**, 257–260 (1990).
- Greenwood, P. F., Strachan, M. G., Willet, G. D. & Wilson, M. A. *Org. Mass Spec.* **25**, 353–362 (1990).
- Waugh, A. B. & Bowling, K. McG. *Fuel Proc. Technol.* **9**, 217–233 (1984).
- Bethune, D. S. *et al.* *Chem. Phys. Lett.* **179**, 181–186 (1991).
- Brooks, J. D. & Taylor, G. H. in *Chemistry and Physics of Carbon* (ed. Walker, P. H. Jr) 282 (Dekker, New York, 1968).

taken in with the food.

There is no information yet on the frequency of shedding, nor on the way fluoride accumulates in the stomach wall. Contamination cannot be the only explanation. How should we continue our observations on our next visit to the Antarctic?

ALBERT J. BEINTEMA

Research Institute for
Nature Management,
PO Box 9201,
6800 BH Arnhem,
The Netherlands

Whaling in the dark

SIR — As part of a comprehensive assessment of whale stocks being undertaken by the International Whaling Commission, I and others have been investigating the genetics of the minke whale (*Balaenoptera acutorostrata*). Preliminary results suggest that minke whales from separate genetic stocks mix in the areas where they are hunted, which complicates the delineation of static boundaries necessary for management. (See the International Whaling Commission's special issue number 13).

Although only two Antarctic management areas (IV and V) were investigated, analysis of mitochondrial DNA, nuclear DNA and isozymes gave consistent results. Whales in the two areas were genetically indistinguishable, but more variable than lower-latitude populations in the North Pacific and North Atlantic, leaving open the possibility that more than one stock was represented in the Antarctic sample. The genetic distance between populations in the North Pacific, North Atlantic and Antarctic (sometimes recognized as three separate subspecies) was as great as the genetic distance between some species in the same genus. A population-level genetic distance was recognized between minke whales caught in the Korean and Japanese coastal fisheries.

At the 1991 scientific committee meeting of the International Whaling Commission further data were presented by S. Wada indicating that whales from the two genetic stocks on either side of Japan are mixing in summer months on feeding grounds north of Japan. Variation within all Northern Hemisphere populations was low for mitochondrial genetic characters, and in some cases lower than isozyme variation, suggesting the possibility of transient population bottlenecks.

Minke whales are typically hunted in areas where they congregate to feed. So far, the genetic data suggest that populations are highly structured, with recognizable genetic stock distinctions over short geographical distances (from one side of

Japan to the other, for example), low levels of variation within some stocks, and that stocks may mix on hunting grounds. However, we do not know how many stocks there are or the degree and pattern of mixing in temporary assemblages. Even though local populations may be large, as in the Antarctic, hunting 'blind' to the pattern of mixing genetic stocks could lead to the selective depletion of genetic variation.

Reduced genetic variation can compromise the viability of a species. Genetic variation in areas of potential breeding (probably in low and middle latitudes) needs to be investigated more extensively and compared with variation in areas where these populations may mix while feeding. Skin biopsies can be collected for these analyses without harming whales or disrupting populations.

A. R. HOELZEL

Department of Genetics,
University of Cambridge,
Cambridge CB2 3EH, UK

(Also at IRC, Population Biology Centre, Ascot, UK.)

A plasma cloud, not a planet?

SIR — The report two weeks ago (M. Bailes *et al.* *Nature* **352**, 311–315; 1991) of periodic changes in the arrival times of signals from the pulsar 1829-10, interpreted by its authors as evidence of an orbiting planet, contains two rather striking coincidences which suggest an alternative class of models for this unexpected phenomenon. The more obvious coincidence is the agreement to less than one per cent (well within the observational uncertainties) of the orbit period with one half of an Earth year. The second coincidence, also accurate to one per cent, is between the Earth's orbital phase and that of the putative planet: when the planet and pulsar are at quadrature, corresponding to zero time-delay in the pulsar's signals, the Earth–Sun–pulsar angle is nearly 90°. These two commensurabilities of the Earth and the pulsar planet orbits allow for a local explanation of the timing observations.

The interposition along the line of sight to the pulsar of a dispersive or refractive structure whose parallax is observable from the Earth could naturally explain the six-month period. The results of Bailes *et al.* provide several immediate constraints on this supposed object. Its angular size must be $\leq 1^\circ$ for its effects to be unobservable in PSR 1829-08, which is only 2° away in the sky. In addition, the object's parallax must be great enough to produce the observed variation in signal transit time between the pulsar and the Earth, but its proper motion must be small enough to sustain the same effect

for the five years covered by the observations.

The peak-to-peak amplitude of the timing residuals is 15 ms. At an observing frequency of 1,400 MHz, this corresponds to a change in dispersion measure of $0.7 \text{ cm}^{-3} \text{ pc}$, which amounts to a 0.15 per cent variation in the total dispersion measure to the pulsar. For a spherical blob 1 AU across, the required change in electron density would be 10^5 cm^{-3} . These parameters are similar to those inferred for the structures that produce the extreme focusing event in extragalactic radio sources. Alternatively, the object could be non-dispersive, but have an index of refraction varying on the same scale; this would provide an oscillation in the pulsar's apparent position in the sky which would, on application of the fitting procedure, generate the observed residual.

To produce the observed timing variations, the object must have structure on a scale D , comparable to the distance between extreme positions of the line of sight of the pulsar: $D < 1 \text{ AU} \times (\text{blob-pulsar distance})/(\text{Earth-pulsar distance})$. But the proper motion v_p must be small enough that the object has not significantly moved across the pulsar line of sight in 5 years: $v_p < D/(5 \text{ yr})$. If the object is in the solar neighbourhood, these arguments require $D = 1 \text{ AU}$ and $v_p < 1 \text{ km s}^{-1}$. This is already a low transverse velocity, and would have to be smaller if the object were farther away. The restriction that its angular size must be $\leq 2^\circ$ means that it would have to be at least 30 AU away; an object orbiting the Sun at several hundred AU would have a sufficiently low transverse velocity. There could be up to 10^5 blobs at such distances without violating the constraint that only one pulsar out of 500 shows an effect.

This class of models has the great virtue of being immediately testable. Simultaneous observation at two frequencies could easily identify changes in the signal dispersion. Monitoring by the Very Large Array over a few months could probably detect changes in the pulsar's apparent position caused by a refractive medium. While we do not regard these physical models as inherently compelling, the identification of the pulsar's putative planet with the Earth effects a certain economy of coincidences, avoids substantial revision of our notions of supernova explosions and planet formation, and admits of straightforward tests that could falsify it forthwith.

DAVID J. HELFAND
THOMAS T. HAMILTON

Departments of Astronomy and Physics,
Columbia University,
New York,
New York 10027, USA

Free lunch?

SIR — The comprehensive review of the patterns and properties of food webs by Pimm *et al.*¹ has been made possible by the recent compilation of large amounts of data on trophic relations in communities². But several unrealistic assumptions widespread in the literature block further insight into the principles of community energetics.

Pimm *et al.* follow ecological tradition and describe food webs as having a lower (or producer) level and an uppermost (or top predator) level. That there are producer organisms is a realistic postulate, although producers can be thought of simply as predators for sunlight and inorganic nutrients. That there are top predators, however, is not tenable, if by top predators we mean "species on which nothing else in the web feeds (page 670 of ref. 1). When a lion or an eagle dies, the matter of which it is composed (and the energy contained therein) does not simply disappear into the ether, nor does it pile up indefinitely, forming vast accumulations of predators' carcasses. Rather, these organisms are consumed by scavengers and detritivores. There is a general perception that these scavengers should not be included in food webs and trophic pyramids because they do not capture live organisms for consumption the way herbivores and predators do. But from the perspective of energy flow, the distinction between species that capture a prey item alive, kill it and then eat it, and species that consume organisms that are found dead is entirely artificial.

Food webs that do not draw artificial distinctions between the so-called top predators and the organisms that feed on them may affect our understanding of trophic patterns in several ways. For example, contrary to the claims of Pimm *et al.*, trophic cycles are probably ubiquitous in food webs. When a top predator, such as an eagle, dies it may be eaten by scavenging flies, which might be fed on by insectivorous birds, which in turn may fall prey to an eagle, forming a cycle. In their Fig. 1, Pimm *et al.* depict various predators, parasites and hyperparasites as not having any species exploiting them; this is probably more a reflection of our ignorance of the natural history of many of these species than it is a biological fact. A more realistic representation of this figure would probably include arrows circling back into the web, forming trophic cycles — especially towards the upper parts of the food web. Cycles are probably not rare, but are perceived as such because of the unfortunate custom of leaving scavengers and detritivores out of the picture, and because of the scarcity of information about what feeds on

the smaller organisms and on the rare predators and parasites. A more detailed knowledge of the natural history of biological communities may reveal that there are few, if any, species that are top predators in the sense of being consumers that do not become, in turn, any other species' lunch.

JOHN E. TOBIN

*Museum of Comparative Zoology,
Harvard University,
Cambridge,
Massachusetts 02138, USA*

1. Pimm, S. L., Lawton, J. H. & Cohen, J. E. *Nature* **350**, 669–674 (1991).
2. Cohen, J. E., Briand, F. & Newman, C. M. *Community Food Webs: Data and Theory* (Springer, New York, 1990).

SIR — Pimm *et al.*¹ discuss community food web theory, based on the results of a previous survey involving analysis of 113 published studies². But only 23% of these studies are from terrestrial ecosystems, the rest being mainly from submerged and semi-submerged habitats. Tropical rainforest ecosystems, containing most of the world's terrestrial species, are represented by only one study, previously described as incomplete³. Therefore, many of the general features of community food webs outlined by Pimm *et al.* may be more applicable to aquatic ecosystems than to many terrestrial ones.

The terrestrial detritus food web is a major component of most terrestrial ecosystems, and is directly responsible for regulating nutrient cycling, litter decomposition and energy flow. None of the 113 studies considered this food web in detail. Although terrestrial detritus food webs are relatively poorly understood, many of the features of food webs outlined by Pimm *et al.* do not seem to conform with the results of some recent studies on terrestrial detritus-based systems.

The following examples demonstrate this: (1) Omnivory is probably widespread in most detritus food webs. This does not apply only to organisms feeding on necromass (as noted by Pimm *et al.*), but also to several below-ground mite and nematode species which can feed on live organisms of more than one trophic level^{4,5}. (2) Within-habitat compartmentalization is a feature common to most detritus food webs. The two main saprophytic groups of organisms, bacteria and fungi, are morphologically quite distinct. Compartmentalization arises because most microbivores are either spe-

cialized feeders on bacteria (protozoa and bacterivorous nematodes) or fungi (fungivorous mites and nematodes)^{6,7}. There appears to be relatively little overlap between these compartments. (3) Many detritus food webs often contain more than four trophic levels⁷, probably as a result of the soil matrix serving as a highly buffered three-dimensional system.

D. A. WARDLE

*Ruakura Agricultural Centre,
Private Bag,
Hamilton,
New Zealand*

Thin-film YBCO magnetometer

SIR — There has recently been much progress in the development of circuits made from thin films of high-transition-temperature (T_c) superconductors. As a result, one can now construct sensitive magnetometers consisting of a d.c. SQUID (superconducting quantum interference device) coupled to a flux transformer to enhance its sensitivity to magnetic fields¹. We have made² a thin-film $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ (YBCO) magneto-

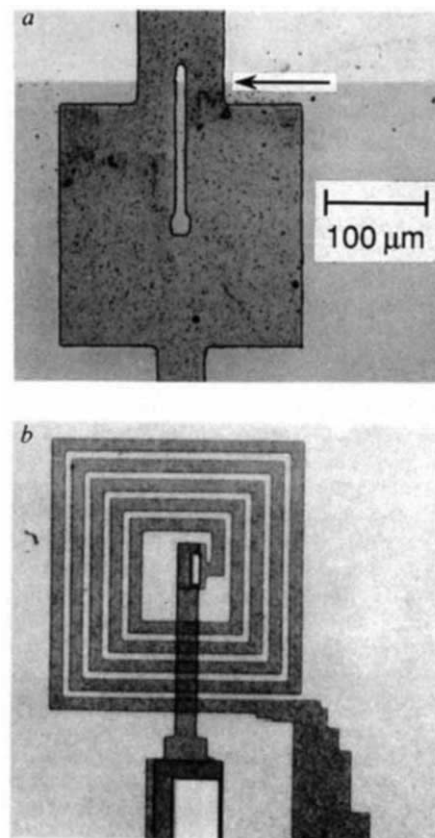


FIG. 1a, YBCO d.c. SQUID. The two Josephson junctions, formed by a 45° grain boundary, are indicated with an arrow. b, Input coil of YBCO flux transformer. The two leads to the input coil are connected to a thin film pickup loop (not shown) with an area of 81 mm².

1. Pimm, S. L., Lawton, J. H. & Cohen, J. E. *Nature* **350**, 669–674 (1991).
2. Cohen, J. E., Briand, F. & Newman, C. M. *Community Food Webs: Data and Theory* (Springer, New York, 1990).
3. Schoener, T. W. *Ecology* **70**, 1559–1589 (1989).
4. Walter, D. E. *Ecology* **68**, 226–229 (1987).
5. Walter, D. E., Hunt, H. W. & Elliott, E. T. *Pedobiology* **31**, 247–260 (1988).
6. Hendrix, P. F. *et al. Bioscience* **36**, 374–380 (1986).
7. Ingham, E. R. *et al. J. appl. Ecol.* **23**, 597–614 (1986).

meter, operating in liquid nitrogen, in which the transformer produced a magnetic-field gain of approximately 80. To demonstrate the sensitivity and practical applicability of the device, we report here the measurement of a human magnetocardiogram. Although a magnetocardiogram has previously been obtained with a bulk YBCO SQUID³, we believe that our experiment is the first to apply a thin-film high- T_c magnetometer to a practical measurement.

Our magnetometer² consists of one chip bearing a d.c. SQUID which is pressed face-to-face with a second chip bearing the superconducting flux transformer (Fig. 1). The SQUID was made using a bi-epitaxial process⁴ by which single 45° grain boundaries can be grown in an epitaxial YBCO film; suitably patterned, the grain boundaries form the Josephson junctions of the SQUID. The SQUID is magnetically coupled to the five-turn input coil of the flux transformer, made from three epitaxial films. The lower YBCO film is patterned into a cross-under line which connects the inner turn of the coil to the pickup loop. This line is insulated from the upper YBCO film, which forms the coil and the pickup loop, by a layer of SrTiO₃. At the ends of the cross-under, windows are etched in the SrTiO₃ to provide contact between the YBCO layers. A magnetic field applied to the pickup loop induces a supercurrent in the transformer and hence a magnetic flux in the SQUID, thereby increasing the effective area and magnetic-field sensitivity of the magnetometer.

We immersed the magnetometer in liquid nitrogen contained in a thin-walled glass dewar extracted from a household thermos flask. The flask was rigidly supported in our shielded room, so that the plane of the magnetometer was parallel to the chest of the standing subject, and within about 25 mm. The dewar and subject were surrounded by two concentric high-permeability cylinders (kindly loaned by Professor Eugene Commins) which attenuated the ambient magnetic fields by a factor of 300. In the absence of any subject, the SQUID exhibited low frequency noise limiting the resolution of the magnetometer to 2.3 pT Hz^{-1/2} at 1 Hz, falling off as $1/f^{1/2}$ (f is frequency) to a white-noise level of 0.35 pT Hz^{-1/2}.

We measured magnetocardiograms from three healthy male subjects ranging in age from 27 to 32. Each subject removed metallic objects from his per-

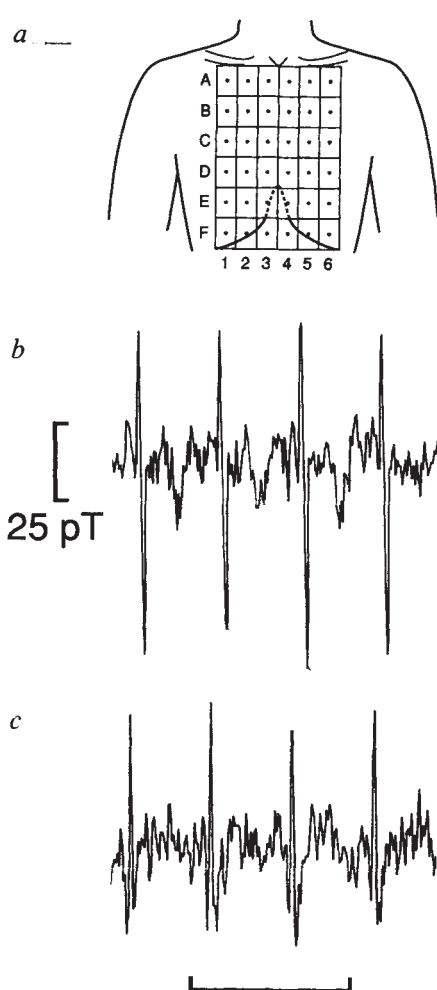


FIG. 2a, Rectangular reference grid on front of chest (redrawn from ref. 5); b, c, magnetocardiograms obtained from same subject at D3 and D5. Scale bars: vertical, 25 pT; horizontal, 1 s.

son, and was not in contact with the dewar. The waveforms were obtained at grid locations⁵ D3 and D5, over the lower-right and lower-left chest of one subject (Fig. 2a). The bandwidth was 2–50 Hz; residual 60 Hz pickup is still evident in both traces. The two cardiograms differ significantly from each other, but each is similar to that previously seen by others⁵ using low- T_c SQUID magnetometers placed at the same grid location.

This result demonstrates that the resolution of thin-film high- T_c magnetometers has attained the level required for practical applications. We note that in a more ideal environment² our magnetometer has achieved a resolution of $(1/f^{1/2})$ pT Hz^{-1/2}, operating at 68 K. For clinical use, an order of magnitude improvement would be desirable. This factor should be achievable with the existing circuit process, which should also allow construction of first-derivative gradiometers that are preferable in some applications. In addition, in the near future, one might also hope for at least

modest reductions in the $1/f$ noise of the SQUID.

A. H. MIKLICH
F. C. WELLSTOOD*
J. J. KINGSTON
JOHN CLARKE

Department of Physics,
University of California and Materials
Science Division,
Lawrence Berkeley Laboratory,
Berkeley,
California 94720, USA

M. S. COLCLOUGH
K. CHAR
G. ZAHARCHUK

Conductus, Inc.,
969 W Maude Avenue,
Sunnyvale,
California 94086, USA

*Present address: Center for Superconductivity Research,
Department of Physics, University of Maryland, College
Park, Maryland 20742, USA.

Metamorphic fluid flow

SIR — From a study of the Connemara Marble Formation at Glencoghan, western Ireland, Yardley *et al.*¹ conclude that the flow of metamorphic fluid during D3 was controlled by local D3 fold geometry, and that loss of this fluid caused the folding to cease. But field and petrographic data suggest instead that there may have been a large-scale circulation of fluid across lithological boundaries.

The rocks studied by Yardley *et al.*¹, diopside gneiss and green marble, occur throughout the Connemara marble Formation outcrop² and form thick units on the limbs of, and folded around, D3 antiforms and synforms. The formation was folded (D2), internally strained, and boudinaged before the D3 folding, and diopside was stable during the D2 metamorphism. Thus it cannot be assumed that the variations in thickness of the metasomatized rocks (A, B in the figure) around the closure of the Glencoghan fold are primary¹, nor that the diopside (or forsterite) which they contain all grew during D3. Also, a thick unit of largely diopside gneiss (C) appears to connect with layer A studied by Yardley *et al.* Unless it can be shown that all of the other diopside gneiss units in the formation (with zoned diopsides up to 4 cm long) apart from layer A formed as a result of isochemical metamorphism not accompanied by infiltration metasomatism, then the thesis that metasomatism in unit A was structurally controlled (D3), and linked with that in B, collapses.

A crucial feature of the hypothesis of Yardley *et al.* is that the fluid responsible for the metasomatism originated from pelites of the Barnanoraun Forma-

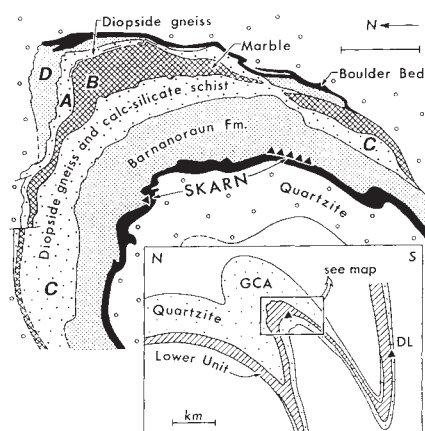
1. Clarke, J. *Proc. IEEE* **77**, 1208 (1989).

2. Miklich, A. H., *et al. Appl. Phys. Lett.* (in the press).

3. Likhachev, A. G. *et al. Supercond. Sci. Tech.* **3**, 148 (1990).

4. Char, K. *et al. Appl. Phys. Lett.* (in the press).

5. Williamson, S. J. & Kaufman, L. J. *Magnetism Magnet. Mat.* **22**, 129 (1981).



Outline geological map of Glencoaghan. Inset shows a vertical, true-scale N-S cross-section through the area. Scale bars: main figure, 250 m; inset, 1 km. GCA, Glencoaghan antiform; DL, Derryclare Lough; Δ , skarn.

tion inside the carapace of Bennabeola quartzite which encloses the older rocks (see inset of figure). However, mineral assemblages in these rocks, including those from the D3 sillimanite grade overprint, show that little or no muscovite breakdown has occurred and (apart from destruction of staurolite) that they did not provide a major source of metamorphic fluid during D3. Also, the pelite band regarded as forming a pathway for this fluid is isolated in two-dimensional section (D) and could not have effectively tapped fluid from the rest of the formation as postulated. Yardley *et al.* claim that bodies of garnet skarn mark high-level, pipe-like conduits by which the metamorphic fluid left the formation and passed into the overlying rocks, but this interpretation is not borne out by their structural location. The garnet skarn at Glencoaghan is a stratiform body located in the inner arc of the antiform; the one at Derryclare Lough is on the limb of a D3 fold; and the skarn near Maam occurs in a D4 culmination but its structural position with respect to any D3 fold is unknown.

Finally, the Glencoaghan antiform is no less tight than other D3 folds in Connemara; it has a broad rounded hinge zone at Glencoaghan because it has folded a kilometre-thick, competent band of quartzite at that locality, and there is no need to invoke local rheological changes in the Connemara Marble Formation due to sudden fluid loss to explain the fold geometry.

P. W. GEOFF TANNER
Department of Geology & Applied
Geology,
University of Glasgow,
Glasgow G12 8QQ, UK

YARDLEY *ET AL.* REPLY — The differences between our interpretation of

regional fluid flow through the Connemara marble and Tanner's arise through differences in our maps; oversimplifications that inevitably accompany a short article; and because of differences in interpretation of the relative ages of mineral growth and deformation. Tanner and Shackleton² carried out the regional survey of this part of Connemara, but our large-scale mapping, on which our cross-section was based¹, allowed us to map out many units not distinguished on Tanner's sketch map, and these provide the detail behind our hypotheses.

Tanner suggests that there are fewer distinct calc-silicate units and that their variations in thickness are due to tectonism rather than primary development. We found that the metasomatic calc-silicate unit or "diopside rock" (Tanner's unit A) is unique and readily distinguished in the field from Tanner's unit C. But a persistent, near monomineralic, diopside rock can only have originated by a metasomatic event when diopside was stable. There is a thin but persistent marble below and to the west of C, not mapped by Tanner, and we consider this to be equivalent to his marble unit B, thinned on the inverted limb of the major D2 fold. Our section¹ shows the thickness variations of both the diopside rock and a green marble not distinguished on Tanner's map. It is clear that the green marble in particular simply does not occur away from the D3 fold crest, but other rocks, including very ductile marbles, show no such tendency to concentrate in the D3 hinge. This does not agree with a tectonic thickening model. The thickness of the diopside rock may vary as a result of D3 strain, but if so this strain must predate metasomatism and diopside growth, because the coarse, random texture of the rock is similar irrespective of whether the unit is attenuated (about 1 m) or thickened (3–4 m).

At a larger scale, Tanner points out that some developments of green marble and calc-silicate are apparently on D3 fold limbs. Here we defer to his more extensive mapping experience, but we make two points. First, his regional cross-section is simplified to show only those fold closures which structural geologists consider to be major ones. On the ground, D3 folding is more pervasive, as is apparent from our section. The marble quarry at Derryclare lough contains abundant small-scale D3 folds, suggesting that it lies in a D3 monofold of antiform-synform pair which may have been on a sufficient scale to be significant for migrating fluids, if not for subsequent geologists. Second, a D2 hinge line in a D3 fold limb would equally act as a linear focus for flow.

Tanner claims that diopside was stable during the D2 deformation as well as during D3. If this were true of diopside in metasomatic rocks it could invalidate our model. But we have seen no evidence for this, and Tanner provides no details, nor indicates the assemblages in which this supposedly early diopside occurs. It is clear from the textures of pelitic schists elsewhere in Connemara^{3,4} that the regional metamorphic grade during D2 was in the garnet zone, and it would be remarkable for the sort of diopside and forsterite marble assemblages observed in the Connemara marble to develop at such a low temperature.

Tanner also questions whether nearby schists could have provided enough water to cause the metasomatism that we report. He points out that muscovite does remain stable at Glencoaghan, and it is more likely that dehydration of staurolite, with muscovite and quartz, was the likely source of the abundant sillimanite. We concur, and in this case we have probably underestimated the volume of source schist required; but as sillimanite schists are common and metasomatised marbles rare in the area, this merely implies a somewhat larger fluid source region than we suggested¹. A further point of disagreement is the extent of the schists overlying the marble sequence, and separating it from a thick mass of quartzite. This is a detail of mapping in a very poorly exposed part of the area, and we see no reason to revise our map. We reiterate that isotope evidence provides a definitive link between pelite and diopside rock.

Tanner also notes that the small, apparently strata-bound skarn occurrences at Glencoaghan underlie the calc-silicates, (although they may well be above similar rocks in deeper parts of the fold stack). However the largest skarn body, which forms the basis for our model and occurs about 16 km to the east of Glencoaghan, has clearly cross-cutting relationships to the host rocks (boulder bed) and structurally overlies the Connemara Marble Formation, which does not in fact surface where the skarn pipe outcrops³.

BRUCE YARDLEY
SIMON H. BOTTRELL
R. A. CLIFF

Department of Earth Sciences,
University of Leeds,
Leeds LS2 9JT, UK

1. Yardley, B. W. D., Bottrell, S. H. & Cliff, R. A. *Nature* **349**, 151–154 (1991).
2. Tanner, P. W. G. & Shackleton, R. A. in *The Caledonides of the British Isles—Reviewed* (eds Harris, A. L. Holland, C. H., Leake, B. E. 243–256 (The Geological Society, London, 1979).
3. Yardley, B. W. D. *J. geol. Soc. Lond.* **132**, 521–542 (1976).
4. Yardley, B. W. D., Leake, B. E. & Farrow, C. M. *J. Petrol.* **21**, 365–399 (1980).

The God of the Galápagos

David L. Hull

Darwin on Trial. By Phillip E. Johnson. *Regnery Gateway, 1130 17th Street, NW Washington DC 20036, USA: 1991. Pp.195. \$19.95.*

In any area of science, at any one time, there will be numerous issues about which there is unanimity of opinion, a few mavericks notwithstanding. However, scientists spend most of their time at the frontiers of their discipline where genuine disagreement is commonplace. Periodically, they also have to backtrack to re-evaluate some tenet that they had all taken for granted for years. Evolutionary biologists are no exception. They agree that species evolve, but they disagree both about the evolutionary process and about specific phylogenies. How important is natural selection in contrast to sexual selection, drift, molecular drive, and so on? Is speciation usually punctuational or do species frequently change *en masse* over long periods of time? How close is *Archaeopteryx* to the fork in the phylogenetic tree that led from ancient reptiles to birds on the one hand and present-day reptiles on the other? Are human beings more closely related to chimpanzees than to gorillas or vice versa?

The same divergence of opinion can be found among scientists about proper scientific method. Some voice extremely empirical, inductivist views of science. Science is an accumulation of facts without any recourse to theoretical speculation. Others see a role for hypotheses in science as long as they are testable in a very strict sense. To count at all, data must be totally independent of the hypotheses that they are designed to test. Others see science as a much more complicated affair. The goal of science is to produce reliable knowledge, but no cut-and-dried method exists that can guarantee success. Because scientific knowledge is so interdependent, rarely is evidence totally independent of the hypotheses that it is used to test. As dogmatic as scientists can be at times, they are committed to fallibilism. Not only may they be mistaken, but quite frequently they are.

In *Darwin on Trial*, Phillip Johnson, a former law clerk for Chief Justice Earl Warren and professor of law at the University of California at Berkeley,

exploits the complex, fallibilist character of science in general and evolutionary biology in particular to urge equal time for his religious convictions. He is not a creationist in the sense that he thinks God miraculously created all species in six days some ten thousand years ago, but he does believe that any purely naturalistic explanation of the creation of life on Earth and the emergence of

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

God Creating Adam: might the God implied by evolution be even more terrifying than the one in William Blake's myth-making cameos?

human beings is inadequate. Instead, as a "philosophical theist and a Christian" he believes that God brought all living things into being to further His own purpose, possibly by creation from nothing, possibly in the way that contemporary evolutionists claim.

Johnson has read much of the semi-popular literature and several texts on evolutionary biology, chiefly the works of Stephen Jay Gould, Richard Dawkins and Douglas Futuyma. He runs through the usual objections to darwinian versions of evolutionary theory — the "lack of transitionals in the fossil record, the sudden explosion of complex life forms at the beginning of the Cambrian age, the difficulty of explaining the origin of the genetic code, the limits to change shown by breeding experiments, the 'hopeful monster' controversy, the punctuated equilibrium controversy, or the importance of catastrophic extinctions". In his discussions, he presents the usual caricatures of evolutionary biology, only to take them back after they have had their effect, a ploy that seems to be common in courts of law. An attorney

will go on at some length about the sex habits of a witness only to end with the admission that such issues really do not bear on the crime being alleged. Johnson runs through the familiar objection that survival of the fittest is a tautology, only to conclude that on certain interpretations it is not. Of course, by then, the damage has been done. The reader is left with the impression that this principle is in some sense suspect.

Johnson spends most of his book arguing that the case for evolution is not proven. In fact, it is so shaky he is led to ask why so many people, including experts whose intelligence and intellectual integrity he respects, can be so blind. Even those scientists who are most vocal in their criticisms of current versions of

Bridgeman Art Library

evolutionary theory acknowledge the very tenet that bothers Johnson the most — the belief that species evolve. For example, the French zoologist Pierre Grassé and the German geneticist Richard Goldschmidt object in no uncertain terms to several basic tenets of darwinism, but neither of these men express any doubts that evolution has occurred.

More recently, Colin Patterson caused quite a flap when he presented a paper at the American Museum of Natural History in which he suggested that looking at character distributions without assuming evolution might

be a productive research strategy. But when Johnson pushed him on the issue, he admitted that he continues to accept evolution as the "only conceivable explanation for certain features of the natural world". Creationists repeatedly quote the claims that Karl Popper made at one time in his career that evolutionary theory as set out by the darwinians is not a genuine scientific theory. Johnson acknowledges that later Popper backed away from this position, not because he realized that he was mistaken but because he was "besieged by indignant Darwinist protests". Johnson also implies that Patterson "eventually disavowed the whole business" after he "came under heavy fire from Darwinists". Anyone who knows Sir Karl or Colin Patterson is likely to doubt Johnson's psychological reconstruction. Neither man is easily bullied.

Why are even the scientists who are most sceptical of darwinian versions of evolutionary theory nevertheless forced to accept evolution? Johnson's answer is their commitment to naturalism. Evolution is the only viable naturalistic ex-

planation for the structure of the living world, and scientists are committed to naturalistic explanations. They are willing to entertain naturalistic explanations as well as those set out by darwinians but not reference to miracles, God's plan, or guiding forces. As inadequate as sexual selection may or may not be as an explanation of the peacock's tail, it is preferable to Johnson's explanation in terms of a "whimsical Creator".

Johnson finds the commitment of scientists to totally naturalistic explanations dogmatic and close-minded, but scientists have no choice. Once they allow reference to God or miraculous forces to explain the first origin of life or the evolution of the human species, they have no way of limiting this sort of explanation. Why does the Earth have a magnetic field, why do organisms use only *laevo* amino acids, why is the savings and loan industry in such trouble? It is easy enough to answer that these phenomena are all part of God's great plan, but in the absence of some partially independent knowledge of God and His intentions, such explanations are no less vacuous than the usual parodies of the principle of survival of the fittest.

The compromise that scientists and theologians have hammered out through the years is that science and religion, when properly construed, cannot conflict. However, this compromise works only if neither side pushes too hard. Advocates of evolutionary ethics transgress this boundary from one side; natural theologians from the other. Advocates of evolutionary ethics claim to provide totally naturalistic explanations of ethics, whereas natural theologians acknowledge inferential relations between God and His handiwork. If the Universe is a perfectly running clock, one sort of God is implied. If, at bottom, natural processes are indeterministic and the basic regularities in the Universe are a function of the contingencies of its origin, quite another God is implied. The problem that biological evolution poses for natural theologians is the sort of God that a darwinian version of evolution implies.

What kind of God can one infer from the sort of phenomena epitomized by the species on Darwin's Galápagos Islands? The evolutionary process is rife with happenstance, contingency, incredible waste, death, pain and horror. Millions of sperm and ova are produced that never unite to form a zygote. Of the millions of zygotes that are produced, only a few ever reach maturity. On current estimates, 95 per cent of the DNA that an organism contains has no function. Certain organic systems are marvels of engineering; others are little more than contraptions. When the eggs that cuckoos lay in the nests of other

birds hatch, the cuckoo chick proceeds to push the eggs of its foster parents out of the nest. The queens of a particular species of parasitic ant have only one remarkable adaptation, a serrated appendage which they use to saw off the head of the host queen. To quote Darwin, "I cannot persuade myself that a beneficent and omnipotent God would have designedly created the *Ichneumonidae* with the express intention of their feeding within the living bodies of caterpillars."

Whatever the God implied by evolutionary theory and the data of natural history may be like, He is not the Protestant God of waste not, want not. He is also not a loving God who cares about His productions. He is not even the awful God portrayed in the book of Job. The God of the Galápagos is careless, wasteful, indifferent, almost diabolical. He is certainly not the sort of God to whom anyone would be inclined to pray.

Johnson dedicates his book in part to "those brave souls" like himself "who

asked the hard questions even when there was never a chance of getting a straight answer; and to those in science who want to allow the questions to be asked." The questions that Johnson asks have been asked over and over again. Most have received very straight answers. Others are still moot. If any scientists have tried to keep these questions from being asked, they have failed miserably. Johnson's problem is that he does not like the answers that he hears. He wants evolutionary biologists to include reference to God in their professional writings in the way that he, I presume, does in his. If Johnson had written a religiously motivated criticism of thermodynamics, quantum theory or plate tectonics, it might have been worth reading, but I cannot imagine why anyone would want to read yet another rehash of creationist objections to evolutionary theory. □

David L. Hull is in the Department of Philosophy, Northwestern University, Evanston, Illinois 60208, USA.

Between the sheets

T. M. Hardman

Paper Chemistry. Edited by J. C. Roberts. Blackie/Chapman and Hall: 1991. Pp.234. £59, \$127.50.

FOR anyone struggling to find a basic text on the chemistry of pulp and paper, this book will come as a great relief. It helps to provide a better understanding of what papermakers call wet-end chemistry, that is, the chemistry of the formation of paper from aqueous suspensions of cellulose fibre and other additives.

Each of the 12 chapters is written by a distinguished author in the field; most of the contributors are affiliated to industry. Two of the chapters are written to provide a background for understanding the other chapters, and it is here that some of the problems arise.

George Roberts concludes that "despite the occurrence of chain folding in cellulose single crystals, most of the available evidence is against such an arrangement in native cellulose". But the dispute over parallel versus antiparallel chain orientation in the structure of natural cellulose is not completely settled. Kurt Meyer originally proposed a model assuming parallel orientation in 1937, but later abandoned it for theoretical reasons. Martin Chang *et al.*, in a more recent review than the references cited by George Roberts, claim that the evidence indicates that antiparallel chain orientation occurs in all cellulose crystals

except, perhaps, vallomic cellulose. That natural cellulose contains crystalline and amorphous regions makes the issue more difficult to resolve.

In the chapter by Tom Lindström, mention is made of how strong and weak acidic groups in sulphite pulps can be determined by conductometric titration and magnesium-elution. But no reference is made to a third way — potentiometric titration — which is referred to in *Papermaking Raw Materials: Their Interaction with the Production Process and their Effect on Paper Properties* (Mechanical Engineering Publications, 1985), and which is being used in the department in which I work. Furthermore, the calculation for cellulose fibres of the zeta potential by electrokinetic methods that is presented by Lindström might raise a few eyebrows, because with these methods the porosity of the fibres would lead to an underestimation.

Nevertheless, this book will have a great deal of appeal to paper scientists and technologists and it is highly recommended as a textbook for students and researchers involved in the chemistry of pulp and paper. □

T. M. Hardman is in the Department of Chemistry, The University, Whiteknights, Reading RG6 2AD, UK.

Correction

Fundamentals of Molecular Evolution by Wen Hsiung-Li and Dan Graur and *Evolution at the Molecular Level* by Robert Selander, Andrew Clark and Thomas Whittam (for joint review see *Nature* 351, 533; 1991) are published in the United States by Sinauer and distributed elsewhere by W.H. Freeman. □

Nobody's fault

Richard Sibson

The San Andreas Fault System, California. Edited by Robert E. Wallace. US Geological Survey Professional Paper 1515. US Government Printing Office, Washington: 1990. Pp.283. \$20.

"THE greatest of faults", said Thomas Carlyle, "is to be conscious of none." For those interested in raising their geological consciousness on active crustal deformation, this publication is most welcome. The San Andreas fault has been a source of fascination to Earth scientists since the great 1906 San Francisco earthquake, the effects of which were well documented in the Lawson report of 1908 (Carnegie Institution of Washington). The pure strike-slip character of the rupture (which was traced for 430 kilometres) and the accompanying geodetic effects led to the recognition of elastic rebound as the cause of shallow earthquakes, with the direct implication that increments of seismic slip reflect long-term motion along the fault zones. Despite this, the existence of a large lateral displacement along the San Andreas fault, one of the key observations leading to the theory of plate tectonics, did not gain widespread acceptance until 1953.

Today, the San Andreas fault system is the most studied active fault system in the world. Research in California has vastly improved our understanding of shallow earthquakes, and the results have been applied worldwide to the engineering of earthquake-resistant structures. But here one has to be a little careful, because in several respects the San Andreas system is atypical. For instance, nowhere else in the world is aseismic slip or fault creep so widespread.

This lavishly illustrated volume is addressed to the Earth scientist who wishes to gain an overview of the San Andreas system, but the lay person and student will also find much of interest. Edited by the doyen of earthquake geologists, it contains ten chapters from fourteen contributors and surveys the geological, geomorphological, geophysical, geodetic and seismic expression of the fault system. As noted in the preface, the book should be treated as an interim report; it documents our present understanding, but also poses questions for research. Undoubtedly, it will be central to graduate seminars on the San Andreas system. Many of the diagrams and photographs are in colour, and all are informative.

Highlights of the book for me include the photographic album of fault features,



The Arashiyama macaques have been intensely studied since their discovery in 1954 on the island of Kyoto, Japan. The original troop rapidly grew in size, splitting in 1966 into two groups, one of which was transferred in 1972 to a ranch in Texas, USA. *The Monkeys of Arashiyama*, edited by L. M. Fedigan and P. J. Asquith, documents 35 years of research on the two sister groups while revealing the similarities and differences between Japanese and Western traditions in primate studies. Published by State University of New York Press, price is \$17.54 (pbk).

and the discussions on Quaternary deformation, seismicity, present-day crustal movements, and the stress/heat-flow paradox. The imperfect nature of the San Andreas transform is emphasized throughout. The boundary between the Pacific and North American plates is not defined by a single fault zone, as many introductory textbooks would have it, but by a complex fault system a hundred or so kilometres wide in which, at different places, the San Andreas fault accommodates from perhaps a half to three-quarters of the total plate boundary motion. Subsidiary vertical motion within the fault system is striking. I recall my first reaction to visiting southern California; what were all these 3–4-kilometre-high mountains doing in an area of strike-slip tectonics? Local irregularities in strike-slip fault systems transfer lateral into vertical motion, giving rise to a furious tempo of geological activity. Basins pull open rapidly and sediments pour in, only for the resulting structures to become deformed and uplifted in their turn as the fault configuration changes.

Of particular interest is the chapter on seismicity, which uses seven years of data from the high-density Californian seismograph network (comprising roughly 64,000 microearthquake locations) to present a detailed three-dimensional picture of earthquake distribution around the transform boundary. The sharp definition of the seismogenic zone (largely restricted to the upper half of the crust) and the structural information contained in the patterns of microseismicity will come as a revelation to most geologists. Also notable is the absence of small earthquakes along the locked segments of the San Andreas

fault that ruptured in the great earthquakes of 1857 and 1906 — a lesson for those who equate areas with low background seismicity with those of low seismic hazard. In developing the network, the US Geological Survey has gained an important tool for investigating actively deforming crust. Its use in interpreting crustal rheology and thermal structure (defining areas of active magmatism, metamorphism and geothermal fluid flux) is far from fully explored.

In any area of active research, it is virtually impossible to publish a large many-authored document of this kind that is fully up to date, but in most respects the editor and authors have succeeded here remarkably well. There is a discussion, for example, of the 1989 magnitude-7.1 Loma Prieta earthquake, which appeared to have ruptured a section of the San Andreas fault that had been predicted to have a high probability of failure on the basis of segmentation analysis. But there are also some notable omissions. Evidence for rotations of crustal blocks about vertical axes within the strike-slip system (for which there is palaeomagnetic evidence) is given scant mention. Recent compilations of marine geological and geophysical data on active faults in the continental borderland off southern California are not fully used. Nor is there any discussion of the detailed structural studies of local transpression and transtension that have been made along the San Andreas fault itself. And although much is made of Tanya Atwater's famous paper, published in 1970, on the plate tectonics of western North America, her recent updated interpretation (*Geological Society of America*, Vol. N; 1989) is not included.

It will dawn on the reader of this book

that in California, as elsewhere along the Pacific rim, seismic risk (the product of hazard and vulnerability) is steadily increasing as population and infrastructure grow. On the Mojave segment of the San Andreas fault due north of Los Angeles, palaeoseismic studies have shown that the average time between major earthquakes over the past seventeen centuries has been about 130 years. The segment last ruptured in the great magnitude-8 earthquake of 1857. The clock is ticking. □

Richard Sibson is in the Department of Geology, University of Otago, PO Box 56, Dunedin, New Zealand.

Driving forces

John de la Mothe

Physics and the Rise of Scientific Research in Canada. By Yves Gingras. McGill-Queens University Press: 1991. Pp.203. Canadian \$37.50, £33.20.

FAR too often, the development of scientific disciplines and institutions is portrayed as being the result of heroic efforts on the part of exceptional individuals. Far too often, the relationship that exists today between teaching and research is thought to be fully mature and at the end of some 'natural' evolutionary process. And far too often, the historical path taken to develop universities and research facilities is thought of in terms of the expression of science's 'will' on its social milieu rather than in terms of a more dynamic interaction between science and its sociopolitical milieu. Happily, in what is a superb example of descriptive ethnographic history of science, Yves Gingras has managed to avoid all of these popular tendencies and has provided readers with a fresh perspective on the transition of physics from a liberal art to a research profession.

To be sure, we are already quite well informed about the institutionalization of scientific research in countries with long-standing and well-established scientific traditions — such as the United Kingdom, the United States, France and Germany. But because of the very recent development of Canada's research system, Gingras's analysis is able to provide us with useful and subtle generalizations about the modern scientific enterprise itself — and not just in Canada.

The conceptual framework used to achieve this proceeds through essentially three phases. The first of these consists of the emergence of a research practice. In countries with rather young scientific traditions, such as Canada or Australia, this most frequently occurs through the

importation of knowledge and techniques (such as through a series of apprenticeships in European laboratories). In countries with older scientific traditions, however, the emergence of new practices — or the development of new disciplines such as physics — can often be explained with reference to the institutionalization of an individual's or group's experiences (witness, for example, the emergence of organic chemistry around Justus von Liebig at Gussien in about 1840).

Once conditions for the emergence of a research practice have been established, the first representatives of this new practice can work to impose a view of the role or function of the university on the university, which, in turn, allows or takes into consideration the new research practices of the group. This promotes the long-term 'reproduction' of both the practice and the discipline. As Gingras argues, it is the second phase — the institutionalization of research — that is critical to the formation of all national scientific communities.

The final phase involves the formation of a social identity which is either based on discipline (for example, the creation of an association) or on profession (for example, through the formation of a corporation). By establishing official representatives for themselves, researchers acquire a social visibility and can thus defend not only their own interests but also those of their discipline. Clearly, the details of Gingras's well-documented and briefly stated framework will be of principal interest to professional historians and sociologists of science. But it is also cast so gently over the bulk of this valuable study that its organizing power is always transparent. It is never obtrusive to the general reader, and it never detracts from the telling of an intricate 'story' about the establishment of research in Canada. Indeed, this story is simultaneously about many things, all of which are important. It is about the transition from teaching to research and research-based teaching; the securing of resources for research; the role of such institutions as the Royal Society of Canada and the National Research Council; the role of such vehicles as the *Canadian Journal of Research*; the importance of other disciplines and professions to the development of new practices; and it is also about the changing status of the research professions *viz* the style of government intervention. All of these stories are deftly told by Gingras.

As he states, "the status accorded research in the present-day university may give the impression that, along with teaching, the production of new knowledge has always been one of the essential functions of this institution." But this

is not the case. During the second half of the nineteenth century, physics teaching in Canadian universities underwent important changes. From natural philosophy, which was taught to all students in the arts faculties and which required only the services of a professor trained in the traditional bachelor of arts model (which was then dominated by philosophy, Latin and classics), Gingras has documented the move to establish fully-fledged departments of physics where students could take temporary leave of their textbooks and be introduced to the techniques of early experimental physics. The major cause lying behind this transformation was the nineteenth-century demand for engineering education created by Canada's industrial development. The institutionalization of engineering education, as well as the development of medical faculties during the 1870s, thus created a growing need for science courses. The social outgrowth in demand directly supported the development of physics departments (and science departments in general) from the 1880s onwards. By the 1890s, a coterie of trained researchers had begun to appear and to find positions in Ontario, Quebec and across Canada, which in turn allowed them — and the country — to be well prepared to contribute to the technological demands arising from the First World War. The external conditions fostered or at least channelled by the newly formed National Research Council, and coupled with the growing social importance of physics, helped to put in place — by 1930 — a stable national system of research in Canada.

Since that time, the transition of Canada's scientific capabilities has continued, shaped by the same forces that brought it into being in the first place. The stresses of interdisciplinary rivalry (for example between physics and engineering) for labour market share, of publication and of research funding in a steady state all continue to resonate even after the period of Gingras's historical analysis has ended. Thus the reader of this book is left with a tacit acceptance of the Gingras framework. Indeed, the reader is left with an articulate and broadly appealing book, which not only provides a general analysis of the driving forces behind the development of professional scientific research, but which also provides both a basis for future comparative studies as well as an extension into an era of a new profession — that of science policy itself. □

John de la Mothe is at the J. F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts 02138, USA and is a science adviser to the Science Council of Canada, 100 Metcalfe Street, Ottawa, Ontario K1P 5M1, Canada.

Making Mars habitable

Christopher P. McKay, Owen B. Toon & James F. Kasting

Mars is believed to be lifeless, but it may be possible to transform it into a planet suitable for habitation by plants, and conceivably humans. The success of such an enterprise would depend on the abundance, distribution and form of materials on the planet that could provide carbon dioxide, water and nitrogen.

It is becoming increasingly clear that humans can alter the environment on a planetary scale. Our introduction of greenhouse gases into Earth's atmosphere at rates sufficient to modify the climate has led to suggestions that we should take active global measures to counteract the predicted warming^{1,2}—in effect, 'terraforming' the Earth. Renewed interest in human exploration and settlement of Mars by using indigenous martian resources and ecological life-support systems^{3,4} has led to the question of the impact of human activities on Mars and the possibility that Mars could be intentionally altered to make it more habitable.

Although there seems to be no life on Mars today, there is considerable evidence that early in the planet's history, liquid water habitats existed, and conditions may have been suitable for the origin of indigenous life, now long since extinct⁵. If Mars did have an early clement, possibly life-supporting, environment, then it is important to our understanding of planetary evolution to develop models for the fate of the early atmosphere^{6,7} and to consider under what conditions, artificial or natural, this environment could return.

Here we consider the possibility that the atmosphere and climate of Mars could be altered to allow terrestrial life forms, and possibly human beings, to survive on the surface (we assume that Mars is lifeless). Several papers have been published on the technical aspects of planetary-scale alteration of Mars^{8–14}. The philosophical and ethical aspects of the problem have also been considered^{15,16}. Our analysis differs from previous discussions of terraforming Mars in two ways. First, although our analysis is based on simplified models, we have attempted a comprehensive treatment, thus exposing substantial problems that were not obvious from studies that considered limited portions of the problem. We hope our analysis will form an initial step towards organizing studies of terraforming Mars so that a quantitative understanding can be developed. Second, we have limited our consideration to technologies that are not far beyond the current art.

We do not, therefore, consider the fundamental physical parameters of a planet to be susceptible to alteration¹². Thus the orbit, rotation rate, mass and volatile inventory are not, at present, plausible targets for anthropogenic change. It may seem surprising that we cannot import atmospheric volatiles, but this can easily be demonstrated by considering the mass of gas required to produce an atmospheric pressure of just 1 bar on Mars, $\sim 4 \times 10^{15}$ tons. The Space Shuttle can lift a mere 40 tons into low Earth orbit, and even heavy lift launchers under consideration have a proposed capacity of only 140 tons. The possibility of obtaining volatiles by moving large asteroids and comets into collision courses with Mars may not be completely unrealistic, but we do not consider this approach here. It would require more than one million comets of 1 km radius to provide the mass equivalent of a 1-bar atmosphere on Mars. We cannot realistically expect to alter the physical properties of Mars by planetary engineering, but we could change the environmental parameters of the planet¹²: the distribution of volatiles, the surface temperature and pressure, atmospheric composition and opacity, planetary albedo, precipitation and humidity.

Habitability

Mars has a rotation rate similar to that of the Earth, and we assume that its surface gravity (0.38g) would be adequate for long-term biological adaptation. As Mars is 1.52 times as far from the sun as the Earth, the amount of sunlight incident on the planet is only 43% of the terrestrial value. This amount of sunlight is much more than is needed for photosynthesis, so we assume that light will not in itself be a limiting factor.

A key parameter of a habitable planet is its average surface temperature. For Earth, this is 15 °C at present; for Mars it is –60 °C. A habitable planet must have liquid water on the surface, so we assume that the average surface temperature must be 0–30 °C. The lower limit is set by the freezing point of water; the upper limit is arbitrarily chosen but is not relevant to terraforming Mars. Accurate determination of the limits on the mean temperature at which the equatorial regions would be too cold or, alternatively, the polar regions too warm would require a detailed three-dimensional model. The range of temperatures specified here is identical to that used by Dole¹⁷ in an early discussion of habitability.

Plants. We consider first a world suitable for plants and anaerobic microorganisms. The carbon dioxide pressure on Earth today (~ 0.35 mbar) is low enough that certain plants (particularly C-3 plants) are often limited in their ability to obtain CO₂ for photosynthesis. C-3 plants lose the ability for net production at concentrations of 35–45 p.p.m. (ref. 18), whereas C-4 plants continue net photosynthesis down to vanishingly small concentrations. In general, net photosynthesis is depressed as the concentration decreases to less than ~ 0.25 mbar (refs 19, 20). We therefore take 0.15 mbar as the lower limit for CO₂. The same value was used by Lovelock and Whitfield²¹. There is no clear upper limit on CO₂ for plants²²; although concentrations over 1 mbar can have adverse effects on certain species²³, Seckback *et al.*²⁴ reported on photosynthetic algae that thrive under pure CO₂.

Oxygen is not required for anaerobic microorganisms, but some is required by plants for aerobic mitochondrial respiration. Plants actually prefer oxygen levels well below the current value. The requirements vary depending on the plant but, in general, net primary production increases as O₂ is reduced from the ambient level of 210 mbar until a point is reached where the concentration is low enough (~ 20 mbar) to cause metabolic complications²⁰. It may be possible to adapt plants to accept even lower concentrations (~ 1 mbar), because the mitochondrial enzyme that requires oxygen has such a strong affinity for it that it can function with oxygen partial pressures of 0.1 mbar (refs 20, 25). Nitrogen is required for all organisms, and the amount of atmospheric nitrogen must be high enough to allow biological nitrogen fixation. Recent work indicates that bacteria can fix nitrogen at pressures of 10 mbar and less²⁶.

Plants and anaerobic microbes can tolerate total pressures much lower than 1 bar. In addition to a few millibars of N₂ and O₂ and 0.15 mbar of CO₂, the atmosphere need only support the vapour pressure of water at ambient temperatures (6.1 mbar at 0 °C). Thus, pressures as low as 10 mbar may be permissible for a planet suitable for plants and anaerobic microorganisms.

Humans. The minimal tolerable partial pressure of oxygen for humans is ~ 130 mbar over a wide range of oxygen mixing ratios and total pressures²⁷. At oxygen partial pressures above those normal at sea level (210 mbar), effects of oxygen toxicity appear²⁷⁻²⁹. At partial pressures only slightly above normal there are demonstrable symptoms after exposure for 200 hours. Oxygen partial pressures of 500 mbar generate symptoms within 24 hours. Dole¹⁷ also suggested an upper limit of 500 mbar. These data suggest that the upper limit to the oxygen partial pressure is close to the present sea-level value. Long-term exposure to pure oxygen at 345 mbar, as was used during the Apollo missions, seems to be tolerated. But there is a problem with flammability of natural materials at high partial pressures of O_2 (ref. 30), suggesting that we should set the limit at $\sim 50\%$ above the present partial pressure. We therefore take 300 mbar as the upper limit for oxygen. For humans, and presumably other animals, CO_2 becomes toxic to the blood buffer system at ~ 10 mbar (refs 27, 31), which is roughly the partial pressure of CO_2 on Mars today.

An atmosphere suitable for humans to breathe on a long-term basis will probably require a buffer gas to prevent oxygen toxicity and spontaneous combustion. On Earth, of course, the buffer gas is nitrogen, N_2 . The Solar System abundances of the elements suitable for use as buffer gases, normalized to silicon, are: He, 3×10^9 ; Ne, 4×10^6 ; N, 2×10^6 ; Ar, 1×10^5 ; Kr, 45; Xe, 4 (ref. 32). On Mars the atmospheric concentrations of the noble/gases (He, Ne, Ar, Kr and Xe) presumably represent the entire available inventory and are negligible³³. Other light elemental gases (H_2 , F_2 and Cl_2) that could be manufactured on Mars are either explosive or toxic. Carbon dioxide is not suitable, as it is toxic above 10 mbar. We have been unable to identify any more complex gas (such as CH_4 , H_2O , CO , HCN , halogen compounds or SF_6) that would be a suitable buffer gas and that could exist in quantities large enough to provide a partial pressure of several hundred millibars. We therefore think that N_2 is the only candidate. Atmospheric N_2 on Mars is also much too scarce to provide a breathable atmosphere; however, nitrogen may exist in solid reservoirs such as nitrate.

At the temperature of the human body, $37^\circ C$, the vapour pressure of water is ~ 60 mbar, which corresponds to the partial pressure of water vapour in the lungs²⁷. Combined with the normal lung partial pressure of CO_2 (50 mbar) and O_2 (130 mbar), this pressure puts a firm lower limit on the total ambient pressure of a pure O_2 atmosphere of ~ 250 mbar (ref. 27). At pressures below this, humans, and presumably other animals with warm body temperatures, would need pressure

suits to survive. For air-like mixtures, an estimate of the lower limit on pressure can be obtained by considering humans fully adapted to life at high elevations. Billings²⁷ states that long-term habitation at 15,000–20,000 ft (500 mbar) is possible, and below 10,000 ft (700 mbar) there is no significant physiological effect. A minimum atmospheric column mass may, however, be required to provide adequate shielding against cosmic and solar radiation. Because of the lower gravity on Mars, an atmosphere with only 390 mbar total pressure would have the same column mass as the terrestrial atmosphere. In Table 1, we suggest 500 mbar as the lower limit for pressure with an air-like gas. Of this, ~ 300 mbar must be buffer gas, as the upper limit on oxygen is near 200 mbar. The upper limit of total pressure for an atmosphere breathable by humans is determined by the narcotic and toxic effects of inert gases at pressures exceeding ~ 5 bar²⁷.

As summarized in Table 1, we have considered two fundamentally different biospheres on Mars. The first is a world suitable only for plants and anaerobic microorganisms, with an atmosphere composed of almost pure CO_2 along with the small amounts of N_2 and O_2 required for plant metabolism. The second is an atmosphere that would be breathable by humans and other terrestrial animals.

Do suitable climate states exist for Mars?

Having specified the environmental requirements that must be met simultaneously for a planet to be habitable, we now consider whether suitable climate states exist on Mars that meet these requirements.

Plant life. We first consider a CO_2 -dominated atmosphere suitable for supporting plant life. The amount of CO_2 needed is determined by the temperature requirements, not by plant physiology. The temperature that would result from a CO_2 -dominated atmosphere can be estimated from the climate-modelling calculations of Pollack *et al.*⁶. Their results indicate that the pressure of CO_2 needs to be slightly above 2 bar for the mean surface temperature to be above freezing. Other studies have found similar values³³⁻³⁷. Of course, on Mars today the temperature can rise above freezing during the day in the equatorial regions³⁸. It is the need to keep the daily averaged temperature high that requires such large amounts of CO_2 .

In the current martian atmosphere, the mixing ratio of O_2 is 0.13% (ref. 33) and is set by photochemistry³⁹. A thick CO_2 atmosphere (2 bar) on Mars might be similar to some model atmospheres proposed for the early Earth^{40,41}. There, vanishingly low O_2 concentrations are predicted because of the reaction of O_2 with reduced volcanic gases or with reduced minerals at the surface. The interior of Mars is relatively quiescent, however, so the volcanic outgassing rate is probably low. Furthermore, the surface is highly oxidizing and might act as a source for O_2 rather than a sink⁴². Ambient O_2 concentrations could therefore be relatively high, even in the absence of biological production. If the surface pressure were 2 bar and the O_2 mixing ratio were the same as at present, the partial pressure of O_2 would be 2.5 mbar, possibly enough for plant respiration, as discussed above. Much higher O_2 partial pressures could be produced by introducing photosynthetic plants.

Perhaps more critical than the supply of O_2 is the supply of N_2 . To provide enough N_2 for nitrogen fixation and for biomass production, the present atmospheric concentration (0.3 mbar) must be raised to several millibars. The possible, but unconfirmed, source is soil nitrogen. Assuming a few millibars of nitrogen stored as nitrites and nitrates that might be released by soil microorganisms, we conclude that a CO_2 atmosphere on Mars with a pressure of 1–3 bar could provide a habitable environment for plants only.

Human environment. To investigate the question of human habitability, we considered a case in which an Earth-like atmosphere was present on Mars: 200 mbar O_2 and 790 mbar N_2 , with CO_2 set to the upper limit for breathable air (10 mbar) and with water vapour present. Such an atmosphere on Earth would result

TABLE 1 Limits for habitability

Parameter	Limits	Note
Global temperature	0–30 °C	Earth temperature 15 °C
Plants only:		
Total pressure	>10 mbar	Water vapour pressure plus O_2 , N_2 and CO_2
CO_2	>0.15 mbar	Lower limit set by photosynthesis No clear upper limit
N_2	>1–10 mbar	Nitrogen fixation
O_2	>1 mbar	Plant respiration
Human breathable:		
Total pressure		
Pure O_2	>250 mbar	Lung water vapour, CO_2 and O_2
Air mixture	>500 mbar	Based on high elevation
CO_2	<5000 mbar	Buffer gas narcosis
N_2	<10 mbar	Set by toxicity
N_2	>300 mbar	Buffer gas
O_2	>130 mbar	Lower limit set by hypoxia
	<300 mbar	Upper limit set by flammability

in temperatures considerably warmer than at present, as it includes 30 times the present terrestrial CO₂ partial pressure.

Figure 1 shows the thermal infrared energy that would be lost to space from Mars if the surface temperature were arbitrarily set to 15 °C with the nitrogen–oxygen atmosphere described above. To perform these calculations we used the one-dimensional radiative convective model described by Kasting and Ackerman⁴³. The energy emitted from the top of the martian atmosphere is compared in Fig. 1 with the black-body emission at temperatures of 288 K and 212 K. As Mars's effective temperature (set by its distance to the sun) is 212 K, the energy emitted to space must be equal to the total energy of the lower black-body curve to balance the planet's energy budget. As the energy emitted by the surface follows the upper black-body curve, the greenhouse effect of the atmosphere would have to lower the outgoing emission accordingly. The reduction in outgoing radiation provided by this Earth-like atmosphere is insufficient (Fig. 1). To balance the radiation budget the opacity of the atmosphere must be greatly increased, particularly in the window region between 800 and 1,200 cm⁻¹.

If we allow the model to adjust the surface temperature until the atmosphere is in equilibrium, we find that the resulting surface temperature is surprisingly cold, -55 °C (see Fig. 2). This can be understood by considering the greenhouse effect on Mars today which amounts to an increase of ~6 K. Adding N₂ and O₂, which have no infrared absorption, broadens the CO₂ lines and increases the greenhouse effect, but only slightly. By contrast, the greenhouse effect on Earth is ~33 K, despite the lower CO₂ concentration. The larger greenhouse effect on Earth is due to the higher concentration of water vapour in the atmosphere. If Mars could be made warmer, the concentration of water vapour would increase, and the greenhouse effect would be more effective. We conclude that an Earth-like atmosphere on Mars would be far too cold to be habitable.

As one possible way to provide an increased greenhouse effect on Mars, Lovelock and Allaby¹³ suggested the introduction of trace amounts of chlorofluorocarbons (CFCs) into the martian atmosphere. These compounds absorb radiation in the window region where CO₂ and H₂O have little absorption.

Gases of particular interest as greenhouse gases are CF₃Br, C₂F₆, CF₃Cl and CF₂Cl₂. We choose these gases because they are composed of elements that could be found on Mars and have relatively long lifetimes against destruction by solar ultraviolet radiation. Mixing ratios of ~1 part in 10⁹ (1 p.p.b.) of these gases can cause a surface temperature increase of ~0.1 K (ref. 44). This combination of gases does a good job of filling in the window region between 800 and 1,200 cm⁻¹. The first of these, CF₃Br, is used as a refrigerant and as a fire inhibitor. In the Earth's atmosphere CF₃Br has a half-life of over 100 years and is only destroyed in the stratosphere. C₂F₆, CF₃Cl and CF₂Cl₂ are even more stable with lifetimes of >500, 400 and 110 years, respectively⁴⁴. Another possible greenhouse gas considered by Ramanathan *et al.*⁴⁵ is SF₆. This has a long lifetime (>500 years), and is a 'strong' absorber, but we do not have enough detailed information to assess its potential as an artificial greenhouse gas on Mars. None of these gases are toxic at concentrations of parts per million⁴⁵.

The required warming on Mars is ~60 °C, or a factor of 700 larger than the changes considered by Ramanathan *et al.*⁴⁴. It is therefore likely that these CFCs would be needed in concentrations of parts per million or more (1,000 times the concentration considered by Ramanathan *et al.*) to provide the additional greenhouse warming needed to warm the nitrogen–oxygen atmosphere on Mars to a habitable temperature. The radiative transfer methods and band models used by Ramanathan *et al.* cannot be applied with confidence to these greatly increased concentrations. For the purposes of illustration we assume that 10 p.p.m. (0.01 mbar) of these gases would need to be introduced into the martian atmosphere. The total weight of this material would be ~4 × 10¹⁰ metric tons, too much to carry from Earth

as was suggested by Lovelock and Allaby¹³. But if we imagine a continuous production of gases on the surface of Mars at a rate that would offset loss by ultraviolet photolysis, the required production rate of material would be ~10⁸ tons yr⁻¹, if the lifetime of this mix of four CFC gases on Mars is taken to be ~400 years. An upper limit on the required production rate of CFCs can be obtained by noting that such gases would probably destroy any ozone layer on Mars and would therefore be photolysed themselves by photons of wavelength 200–300 nm. If we assume that each of these photons destroys a single CFC molecule, of mean relative molecular mass 200, then ~3 × 10¹² tons of CFCs must be produced each year to offset the loss. Current production of CF₃Cl (CFC-11) and CF₂Cl₂ (CFC-12) on Earth total ~10⁶ tons yr⁻¹ (ref. 46). As discussed later, the concentrations of Cl, F and Br on Mars are probably adequate to provide these quantities of gases.

To assess more quantitatively the effect of adding greenhouse gases to a nitrogen–oxygen atmosphere on Mars, we have modified the radiative–convective model of Kasting and Ackerman⁴³ to include an arbitrarily specified absorber. Lacking adequate information on the optical properties of these CFCs at high column densities, we have considered two cases with a uniformly mixed absorber in our calculations. In one case, the absorption occurs in the window region (800–1,200 cm⁻¹) of the thermal infrared spectrum only. In the other, the absorption covers the entire thermal infrared spectrum with the same opacity at all wavelengths (grey absorption). The surface temperature is plotted in Fig. 2 as a function of the grey opacity added to the model. As can be seen, the first case (absorption in the window region only) does not produce enough warming to raise the planetary average temperature above freezing even when the opacity is high; radiation loss becomes dominated by other spectral regions. In the second case, the temperature rises steeply with absorber amount once the opacity becomes close to unity. Temperatures above freezing are reached when the absorber concentration corresponds to an optical depth in the thermal infrared of ~10. Also shown in Fig. 2 is the effect of adding the absorber active in the window region to the present martian atmosphere rather than the 1-bar nitrogen–oxygen atmosphere.

It therefore seems possible, in principle, to produce a warm oxygen-rich martian atmosphere, given an adequate supply of N₂, O₂ and the appropriate trace greenhouse gases. One of the chief problems is the possibility that the trace gases would prevent the formation of an ozone layer. Then, as discussed

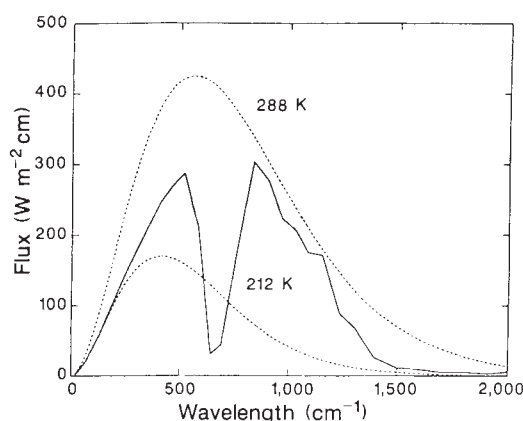


FIG. 1 Thermal balance of a 1-bar nitrogen–oxygen atmosphere on Mars with 10 mbar of CO₂ in equilibrium with water, and a surface temperature set to 15 °C. Shown as the solid line is the thermal infrared radiation emitted from the top of the atmosphere for $T_s = 288$ K. Also shown are the black-body curves at 15 °C and -60 °C (dotted lines). To be in equilibrium, the area under the curve of outgoing infrared flux must equal the area under the curve for the -60 °C black body. The surface is too warm (by ~70 °C) to be in equilibrium.

above, production of these trace gases would have to be at a high enough rate to compensate for their loss by ultraviolet destruction, although they would effectively shield the surface. The possibility of producing a plant-habitable martian atmosphere is higher because of the CO₂ greenhouse effect. We describe next how such a CO₂ greenhouse might be accomplished.

The runaway CO₂ greenhouse effect

Early discussions of warming Mars focused on the possible 'runaway' greenhouse effect of a CO₂ atmosphere in contact with a presumed polar reservoir of CO₂ ice. An initial warming is amplified by sublimation of CO₂, adding to the warming and subliming more CO₂ (refs 8, 47, 48). Carbon dioxide adsorbed into the regolith is another reservoir on Mars that may be susceptible to a runaway effect.

Polar caps. Gierasch and Toon⁴⁷ found that advection of heat from mid-latitudes on Mars becomes an important source for warming the poles when the surface pressure is increased to ~100 mbar. This resulted in a runaway condition, subliming the polar ice caps, if atmospheric pressure were raised above a critical value. These calculations neglected the greenhouse effect entirely. The positive feedback between CO₂ pressure and temperature is much stronger when the greenhouse effect is included. We have extended their model by treating the latitudinal temperature difference as a simple exponential in pressure and by allowing the mean surface temperature to vary with atmospheric CO₂ (ref. 6). We can then express the polar temperature as

$$T_{\text{pole}} = T_{\text{mean}}(\text{CO}_2) - \Delta T e^{-P/P_0} \quad (1)$$

where $T_{\text{mean}}(\text{CO}_2)$ is the mean planetary temperature (Fig. 4a), ΔT is the difference between the mean planetary temperature and polar cap temperature with no advective transport of heat, and $P_0 = 200$ mbar, determined from a fit to the middle curve in Fig. 1 of ref. 47. This expression gives results that are significantly different from the earlier analysis⁴⁸ (Fig. 3). A low-pressure stable state exists, but as the surface pressure increases a transition point is reached (an unstable equilibrium) above which there is no stable state and the polar cap sublims completely. The final temperature and pressure are set by the available CO₂.

This analysis presumes that there is a large permanent reservoir of CO₂ at the martian poles. But the existence of a permanent

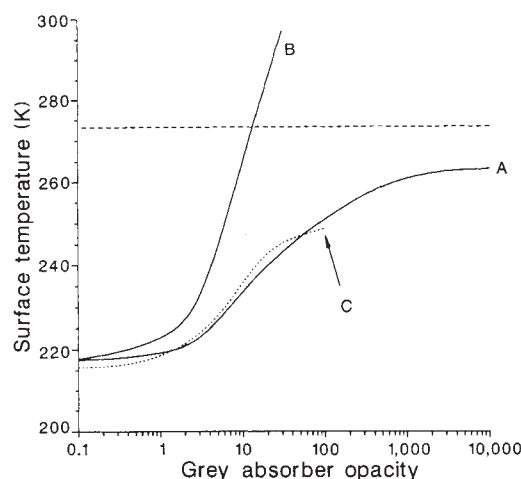


FIG. 2 The effect of adding an absorber on the surface temperature of Mars. For the 1-bar nitrogen-oxygen atmosphere (see Fig. 1), two types of absorbers were considered: curve A is for an absorber that is active in the 'window' region of the thermal infrared spectrum (800–1,200 cm⁻¹) only, curve B for an absorber that is active over the entire spectrum. Curve C shows the effect of adding an absorber active only in the window region to the present 6-mbar CO₂ atmosphere. The dashed line indicates the freezing point of water.

CO₂ cap can be ruled out for the north pole⁴⁹, although there is evidence for a year-round CO₂ reservoir at the south pole⁴⁹. The size of this reservoir is not known, but the residual cap is small, only ~350 km in diameter, and basal melting should limit the thickness of CO₂ it could contain. A 1-km-thick cap of this diameter could provide ~100 mbar of CO₂, and hence a more detailed assessment of the existence and stability of a permanent CO₂ cap at the south pole on Mars would be of interest.

Regolith. Another possible positive feedback results from CO₂ adsorbed into the martian regolith, estimated to be as much as 300 mbar of CO₂ (refs 50–52). To model the desorption of the regolith CO₂ on heating we use the following expression⁵³

$$M_a = C \exp(-T/T_d) P^\gamma \quad (2)$$

where T and P are the surface temperature and pressure, M_a is the total column amount of CO₂ adsorbed and C is a normalization constant that depends on the specific surface area of the adsorbent and the depth of the regolith. C determines the total mass of adsorbed gas. T_d and γ are constants that determine the response of adsorption to temperature and pressure, respectively. Physically, T_d is the temperature change required to outgas a fraction 1/e of the regolith CO₂; small values of T_d indicate that the CO₂ is only weakly bound. As we are interested in the effects of warming the regolith, γ is of secondary interest and we use the value of 0.275 from Toon *et al.*⁵³, which is between the two recommended values from Zent *et al.*⁵⁴. In our calculations, we have let T_d vary from 10 K to 60 K. Estimates from materials analogous to Mars soil suggest $T_d \approx 60$ K (refs 50, 53), but Toon⁵³ noted that this overestimates the mass of adsorbed gas at higher temperatures. Data for adsorption at higher temperatures and pressures are not available for Mars analogue materials, but for activated carbon at a partial pressure of 0.5 bar, $T_d \approx 35$ K (ref. 55). Although the amount of CO₂ indicated for the regolith corresponds to only 300 mbar, we have set it to 1 bar to illustrate a maximal temperature-driven desorption response.

In Fig. 4, we have plotted the partial pressure of CO₂ that would result from heating the regolith for two limiting cases. In the first case, the regolith containing the CO₂ is spread over the entire planet and we treat it as being at the average planetary temperature. In the second case, we assume that only the polar regolith holds any appreciable CO₂. The latter case may be more reasonable, as cold regions will preferentially collect the CO₂. In this case the temperature that governs the release of regolith CO₂ is T_{pole} from equation (1). In all cases the coefficient C is determined by the total amount of CO₂ assumed to be in the atmosphere-regolith systems. The implied depth and specific surface area of the regolith corresponding to the coefficient C depend on the current surface temperature (either the mean surface temperature of 217 K or the polar temperature of 150 K) and on the current CO₂ pressure (6 mbar). For our calculations the total amount of CO₂ is the important quantity because as $T \rightarrow \infty$, all the CO₂ goes into the atmosphere. Figure 4a shows the results for the case of the global regolith, Fig. 4b for the polar regolith. Similar curves are obtained when the total amount of CO₂ is varied except that the pressure asymptote shifts. The results are relatively insensitive to the pressure exponent (γ) over the range 0.2–0.35.

Stable states exist where the desorption curves cross the atmospheric (or polar) temperature curve, and where the slope of the atmospheric curve is less steep than the slope of the desorption curve. For all values of T_d , the current condition is stable in both cases; all curves cross at this point by construction. For the global regolith, the only curve that is of interest is the top one ($T_d = 10$ K). In this case, the desorption curve crosses the atmosphere curve thrice and two stable states exist: one has most of the CO₂ adsorbed into the regolith and the other has most of it in the atmosphere. Unfortunately, a value of $T_d = 10$ K is not easily justifiable. For the polar case the results are more promising: multiple stable states exist for $T_d < 40$ K.

TABLE 2 Total volatile inventory for Mars

	CO ₂ (mbar)	N ₂ (mbar)	H ₂ O (m)	Note
Required for terraforming	≥2,000	≥300	≥500	
Present Mars atmosphere	~10	0.2	~7 × 10 ⁻⁶	Atmosphere only
Earth scaling†	27,000	300	1,200	Equal volatiles (g per g)
Rasool and Le Sargeant ⁵⁸	198	3.1	5.9	³⁶ Ar, ordinary chondrites
Anders and Owen ⁵⁷	140–525	2–8	9.4	K, ⁴⁰ Ar, ³⁶ Ar/ ⁴⁰ Ar
McElroy <i>et al.</i> ⁶⁰	1,760	21.5	133	¹⁴ N/ ¹⁵ N
Clark and Baird ⁵⁹	187–410	8.6	88	⁴⁰ Ar, 'excess volatiles'
Pollack and Black ⁶¹	1,000–3,000	6.6–66	80–160	N, noble gases, Venus data
Carr ⁶⁷	10,000–20,000	100–300	500–1,000	Geomorphology
Dreibus and Wänke ⁶²	(3,000)‡	(33)‡	130	SNC meteorites, martian CI
Greeley ⁶⁶	(>1,000)‡	(>11)‡	>45	Volcanism only

This compilation is adapted from McKay and Stoker⁵.

† Determined by assuming that the martian ratio of volatiles per gram of planet is the same as the Earth's. Earth volatiles⁸³: CO₂, 190 bar, N₂, 2 bar, H₂O, 3,200 m.

‡ These values are not from the referenced source but were determined assuming that the ratios of volatiles are the same as the Earth scaling result.

Thus, an interesting possibility seems to exist: in the case of the polar regolith and $T_d = 20$ K, if the surface temperature could be raised (for example, by adding of greenhouse gases) by ~25 K, raising the surface pressure above the unstable equilibrium point at 30 mbar, the system would switch to the new stable state at a surface pressure of ~800 mbar and a surface temperature of 250 K. For the global regolith ($T_d = 10$ K) a similar transition occurs at 100 mbar, with the final state having essentially totally outgassed the regolith CO₂. Clearly, the existence of stable states which result from the desorption of CO₂ depends on the total amount of CO₂ in the atmosphere-regolith system, the latitudinal distribution of the regolith and, very sensitively, on the parameter T_d . The final temperature of the steady state depends only on the total CO₂ outgassed, and can be determined by the observation that most of the CO₂ is desorbed. Thus, a 2-bar regolith reservoir would have a stable state at ~273 K; a 3-bar reservoir would produce a temperature >280 K.

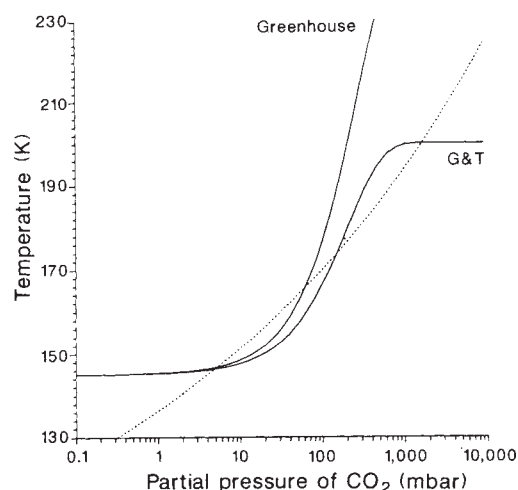


FIG. 3 Equilibrium temperature of a polar CO₂ cap on Mars as a function of surface pressure. The dotted line is the vapour pressure of condensed CO₂. The solid lines are models of the dependence of the polar-cap temperature on the total atmospheric pressure. The curve labelled G&T is a typical result from the work of Gierasch and Toon⁴⁷ who considered only advective transport of heat from mid-latitudes to the poles. For the case represented by this curve, stable states exist at a few millibars and at several hundred millibars total pressure. The other solid curve includes the greenhouse effect, and the polar-cap temperature is now given by equation (1). In this case there is no high-pressure stable state with a persistent polar cap; the cap totally sublimates in a 'runaway' effect.

Here we have assumed large, weakly bound, reservoirs of CO₂ in the martian regolith. The CO₂ released is useful for constructing a plant-habitable atmosphere but may be undesirable in a human-habitable atmosphere. Unfortunately, there are insufficient data to determine the definite role of regolith desorption in warming Mars, but our results suggest that it is an area worth further investigation.

Mars volatile inventory

We now consider whether the volatile inventory on Mars is sufficient to provide the gaseous constituents of a habitable atmosphere. Table 2 shows our estimate of the total inventories of CO₂, N₂ and H₂O that would be needed to construct habitable environments on Mars. The amount of CO₂ is that required to raise the mean temperature to near the freezing point. The amount of N₂ is that required to provide a buffer gas for a breathable atmosphere (see Table 1). The required amount of water is somewhat arbitrary but, as will be discussed below, large bodies of water are probably needed to sequester carbon-bearing sediments, thus liberating O₂. In addition, a habitable planet probably requires an active hydrological cycle. Also listed in Table 2 are various published estimates for the volatile endowment of Mars^{5,7}.

Early theoretical models of Solar System volatiles⁵⁶ predicted that Mars should be more volatile-rich than Earth, as it formed further from the sun. A second geochemical approach^{57–59} is to scale the volatile inventory by the observed argon concentrations, both nonradiogenic ³⁶Ar and radiogenic ⁴⁰Ar. McElroy *et al.*⁶⁰ used the preferential loss of ¹⁵N compared with ¹⁴N to estimate the total nitrogen inventory and then scaled this by Earth's ratios to obtain CO₂ and H₂O. Pollack and Black's method⁶¹ used both the argon and nitrogen values. All of these geochemical models suggest that Mars is volatile-poor compared with the Earth. Dreibus and Wänke⁶² based their estimate of water abundance on an analysis of the SNC meteorites, which are thought to have come from Mars^{63–65}. Greeley⁶⁶ calculated a lower limit by considering only the amount of water that should have been released by the observed volcanic features, assuming 1% as the water content of martian magma.

An alternative method of estimating the amount of water present on early Mars is to use geomorphological evidence. For example, Carr⁶⁷ determined the amount of water that must have flowed through the Valles Marineris system to create the observed channel features. He obtained a lower limit by assuming maximal efficiency for erosion by the flowing water. By extrapolating these results to the rest of the planet, he inferred that Mars has outgassed water equivalent to a depth of 0.5–1 km; this implies that Mars is volatile-rich, with a volatile endowment comparable to the Earth scaling result (Table 2).

All estimates of the initial complement of volatiles are greater than the abundance of water, CO₂ and nitrogen currently residing in the atmosphere of Mars (Table 2). The present loss rate of volatiles due to atmospheric escape is insufficient to deplete the initial endowment of water and other volatiles, so most of the volatiles may still be on the planet in subsurface reservoirs. If the rate of atmospheric escape has remained constant for water (as 2H and O) at $6 \times 10^7 \text{ cm}^{-2} \text{ s}^{-1}$ (refs 60, 68) and for N₂ at $5.6 \times 10^5 \text{ cm}^{-2} \text{ s}^{-1}$ (ref. 69), the total loss over the past 4.5 Gyr would be 2.5 m of water and 1.4 mbar of N₂. That the current D/H ratio on Mars is 6 times as large as on Earth^{70,71} may only reflect fractionation processes in the current atmospheric reservoir⁷². It is also possible that Mars lost a considerable fraction of its early atmosphere as a result of erosion by high-velocity impacts⁷³.

The total complement of volatiles on Mars and their fate over geological time is therefore highly uncertain. If the upper range of estimates is correct, then Mars has the volatiles needed to achieve both habitable states discussed above. If the lower range is correct, or if Mars has lost its initial volatiles to space, then it is unlikely that any habitable state can be produced. This is an important area in which we need better information to evaluate the possibility of terraforming Mars.

If large volatile reservoirs do exist on Mars, they could take several forms. The polar caps may hold large amounts of water ice, perhaps as much as 5,000 km³ (ref. 74), corresponding to a layer of water 4 cm thick covering the entire planet (probably not enough to provide for a green planet). Much larger amounts of water may be tied up as permafrost in the regions poleward of 30° (ref. 75). If there is indeed water equivalent to a depth of hundreds of metres on Mars, this must be where it is located.

As discussed earlier, CO₂ could be stored in the south polar cap (<100 mbar) and in the regolith (~300 mbar). Another, possibly much larger, reservoir for CO₂ is carbonate rocks. Carbonates could have been formed early in Mars's history by reactions of CO₂ with silicate minerals in the presence of liquid water⁶. The existence of a carbonate reservoir on Mars is uncertain, although there is some recent spectral evidence for it⁷⁶. Releasing the CO₂ from carbonate materials would be much more difficult than release from the polar cap or regolith¹⁴.

The most serious potential shortfall of material is N₂ (Table 2). If the present amount of nitrogen in the atmosphere represents the available amount, then this will severely limit the biological possibilities for Mars. Even under the assumption that the C/N ratio of biomass can be stretched to values near 100 (by reduced requirements in nitrogen-limited ecosystems and preferential recycling of N compared with C in sediments), the current N₂ inventory of 0.2 mbar would only allow for the production of 60 g cm⁻² of sediment biomass. This amount of carbon fixation would release only 40 mbar of O₂. Clearly the question of non-atmospheric reservoirs of nitrogen on Mars is a critical, if not the most critical, question that must be answered to assess the feasibility of terraforming Mars. It has been suggested^{5,7} that nitrogen would have been tied up as nitrate on an early wet Mars, and it is possible that considerable nitrogen reservoirs exist in the soil.

In addition to C, H, N and O, the elements S and P are required for life. Sulphur is available on Mars with concentrations reported at the Viking landing sites 10 to 100 times those in terrestrial rocks and soils⁷⁷. Phosphorus was not measured by Viking because its X-ray fluorescence signal was masked by S and Si. Nevertheless it is thought to be present⁷⁸. Thus all the

FIG. 4 Desorption of CO₂ from *a*, a global regolith and *b*, a polar regolith for assumed total CO₂ equivalent to a partial pressure of 1 bar. The four solid curves correspond to values of T_d of 10 K, 20 K, 40 K and 60 K (see equation (2)). The dotted line shows the average temperature on Mars from climate model calculations⁶. In *b*, the dotted line is the polar-cap temperature as a function of pressure, from equation (1). Where a solid line intersects the dotted line, there is a potential steady state.

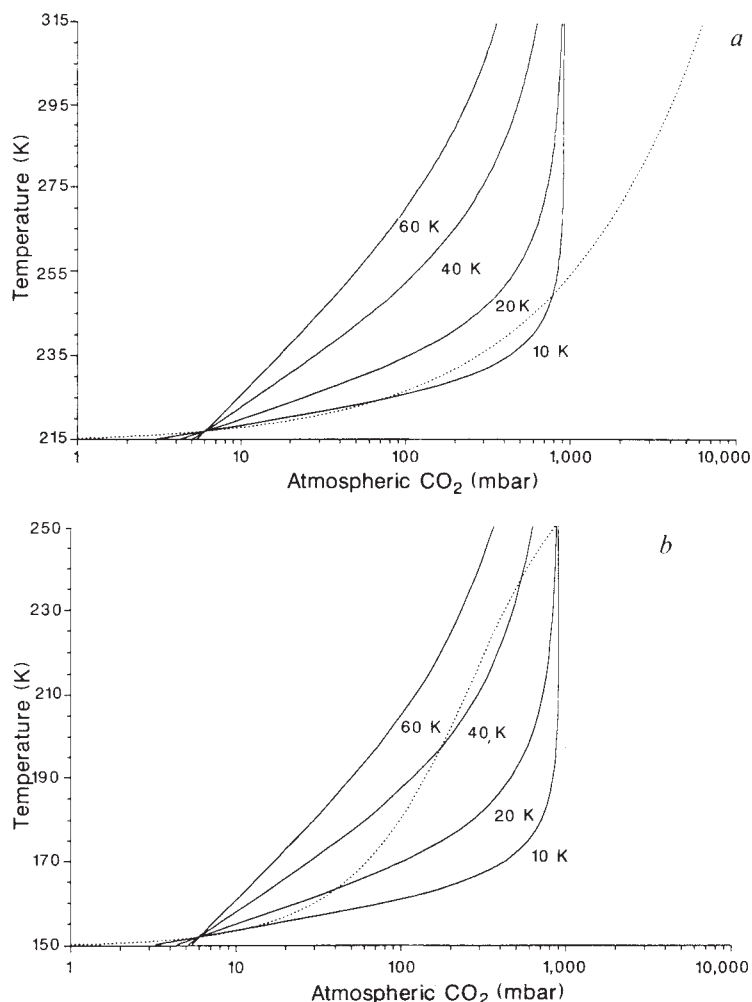


TABLE 3 Energy requirements for terraforming Mars

Initial state	Final state	Amount (g cm ⁻²)	Energy (J cm ⁻²)	Energy/solar power* (years)
CO ₂ (s) at 150 K	CO ₂ (g) at 288 K	2 bar; 5,400 g cm ⁻²	3.7×10^6	7.9
Dirt at 215 K	Dirt at 288 K	~10 m; 2,000 g cm ⁻²	1.2×10^5	0.3
H ₂ O(s) at 215 K	H ₂ O(l) at 288 K	10 m; 1,000 g cm ⁻²	5.5×10^5	1.2
H ₂ O(s) at 215 K	H ₂ O(g) at 288 K	20 mbar; 54 g cm ⁻²	1.6×10^5	0.33
H ₂ O(s) at 215 K	H ₂ O(l) at 288 K	500 m; 50,000 g cm ⁻²	2.8×10^7	56
CO ₂ (g) + H ₂ O	CH ₂ O + O ₂ (g)	200 mbar; 540 g cm ⁻²	8×10^6	17

* The global averaged solar energy on Mars is 4.68×10^5 J cm⁻² yr⁻¹.

major elements of life (C, H, N, O, P and S) are present on the surface of Mars, and most of the trace elements required, such as Fe, Mg and Al, have also been directly detected. To construct the four CFC gases we suggested as greenhouse gases, Cl, F and Br would be required. Chloride has been observed on Mars (0.007 ± 0.003 by mass in the soil⁷⁸). The Solar System ratio of Cl to F (which was not detectable by the Viking landers) based on meteorites³² is 6.2, implying considerable F on Mars (~0.1% by mass⁷⁹). The F concentration is only ~40 p.p.m. in the SNC meteorites thought to represent martian materials⁸⁰, but these also contain only 108 p.p.m. of Cl, well below the Viking value. The availability of F therefore remains uncertain. Bromine was detected in two or three spectra on the Viking X-ray instrument⁸¹ and is estimated to be 20–30 p.p.m. (B. Clark, personal communication). The Solar System abundance of Br³² would imply a concentration in the martian soil of ~2 p.p.m. by mass if scaled to Si, and 16 p.p.m. if scaled to Cl, compared with 0.15 p.p.m. determined from the SNC meteorites⁶².

Timescales for terraforming Mars

We next consider how a suitable steady state might be achieved, assuming that the volatile inventory is adequate to achieve it. From energy considerations, terraforming Mars breaks down naturally into two steps: warming the planet, and altering its chemical state. In Table 3 we have listed the initial and final states associated with this process and the energy required to achieve the change in state. The energy requirements are also listed in terms of the equivalent number of years of sunlight, assuming 100% could be used. This provides a strong lower limit to the time required to achieve the desired transition.

Warming the planet. A plant-habitable environment could be established by exploiting the positive feedback of a CO₂ atmosphere, as discussed above. If the temperature of the martian surface were increased, say, by warming the poles with giant mirrors, or by spreading black soot over the polar caps, or by introducing greenhouse gases, then the amount of CO₂ and water vapour in the atmosphere would also increase. This would have the effect of warming the atmosphere still further and so on. As discussed above, it is possible that above a certain CO₂ pressure the process would accelerate until a high-pressure stable state was reached or until all the CO₂ was outgassed. Regardless of the mechanism, the energy required to complete this step would be ~ 10^6 J cm⁻², equivalent to 10 years of Mars solar energy. This amount of energy corresponds to subliming and warming 2 bar of CO₂, warming a layer of martian regolith 10 m thick, melting a layer of water 10 m thick and evaporating moisture into the atmosphere. Assuming that the process could be sustained and could use 10% of the solar energy on Mars, it would require ~100 years. On a much longer timescale, the deep layers of permafrost ice would melt. To melt a layer of water 500 m deep over the entire planet requires 55 years of solar energy. But the process would probably be limited by the rate of diffusion of heat into the regolith, which might require ~ 10^5 years. If the 2 bar of CO₂ were distributed over a 500-m depth of regolith, it could require ~100 years to diffuse out of the soil if driven by a pressure gradient, but 10^5 years for the soil to be warmed enough at depth to degas it⁵³.

Chemical modification of the atmosphere. The only mechanism that seems capable of changing the bulk composition of a planetary atmosphere is planetary-scale biology, producing O₂ from CO₂. The energy required for O₂ production is larger than that required for the initial warming step. Furthermore, current biological ecosystems have very low efficiencies for conversion of incident solar energy into biomass energy. Net primary production in terrestrial ecosystems ranges from 0.001 to 0.1 g cm⁻² yr⁻¹, with a global average of 0.033 g cm⁻² yr⁻¹ (ref. 82). These numbers correspond to energy efficiencies of 1.5×10^{-5} , 1.5×10^{-3} and 5×10^{-4} , respectively. For a typical average efficiency of (~ 10^{-4}), the production of a breathable atmosphere would require over 100,000 years. Specially developed organisms could yield much higher efficiencies⁸. The limiting step in the production of any net O₂ on Mars may not be the efficiency of photosynthesis, however, but the necessity of sequestering the organic material in places (presumably deep sediments) where it is not reoxidized. This, in turn, probably requires an active hydrological cycle and deep stable basins. Unless vast amounts of water could be liberated, this step could prove to be difficult. This is yet another reason why a human-habitable environment is much more difficult to construct than a plant-habitable one.

The timescales discussed above can be compared with the timescale for the decay of a thick atmosphere on Mars. In the case of a 2-bar CO₂ atmosphere, the dominant loss mechanism would presumably be the formation of carbonate rocks. Pollack *et al.*⁶ have estimated that the lifetime of a thick CO₂ atmosphere on early Mars is ~ 10^7 years in the absence of any recycling mechanisms. Maintaining the atmosphere for longer than 10^7 yr would require some mechanism, either biological or mechanical, for recycling carbonate materials into the atmosphere¹⁴. Nonetheless, it seems that the processes acting to remove the atmosphere of a habitable Mars may be significantly slower than the processes that could create such an atmosphere, so that the resulting system could persist for tens of millions of years.

Conclusions

The complexity of the problem of terraforming a planet does not allow us to draw detailed conclusions. The problem is similar in scope to understanding the Earth's biosphere and biogeochemical cycles. Our work has been a preliminary attempt to identify the important issues and to quantify the problem. Our main deduction is that on Mars a CO₂-rich, plant-habitable atmosphere would seem feasible to construct, given adequate reservoirs of CO₂ and H₂O, whereas an O₂ rich, human-habitable, atmosphere would be very difficult to construct and possibly impractical to warm.

From our analysis, one could propose the following sequence of events: production of CFCs (or other greenhouse gases) starts on Mars and the surface temperature warms up by ~20 K. The regolith and polar caps release their CO₂ and the pressure rises to 100 mbar. One of two things could then happen. If there were large regolith and polar CO₂ reservoirs the pressure would continue to rise on its own. If these were absent, the CO₂ pressure would stabilize, and additional CO₂ would have to be released from carbonate minerals. At this point (perhaps between 100

and 10⁵ years) Mars may be suitable for plants. If there was a mechanism for sequestering the reduced carbon, these plants could slowly transform the CO₂ to produce an O₂-rich atmosphere in perhaps 100,000 years. If sufficient N₂ could also be released from putative soil deposits, and the CO₂ level kept low enough, then a human-breathable atmosphere would be produced. Continued production of CFCs that absorb radiation across the whole spectrum would be required to maintain the warm temperature. Destruction of ozone by these CFCs would probably require these gases be made in sufficient amount (considerably in excess of current terrestrial production rates) to constitute an ultraviolet shield. This proposed process for terraforming Mars relies only on processes that have been demonstrated, and in fact are current, on Earth.

Our analysis is incomplete for several reasons. Most important, our knowledge of Mars is insufficient to determine many key quantities. In addition, we have used only simple models or estimates of many key processes. But our results do suggest that further investigation would be fruitful and that terraforming studies can be made quantitative. We have identified several areas that warrant further investigation: (1) the availability of N₂ and other volatiles on Mars; (2) more detailed models of the processes for warming Mars from its present state with 'runaway' effects involving both the polar caps and regolith; (3)

chemical and biological methods for releasing CO₂ from carbonates; (4) the effects of artificial heating mechanisms, such as albedo changes, greenhouse gases and solar mirrors, on the present martian climate; (5) the photochemistry and O₂ mixing ratio of a 2-bar CO₂ atmosphere on Mars; (6) photochemistry of O₃ in a 1-bar nitrogen-oxygen atmosphere on Mars with parts per million of halogenated gases; (7) climate of a thick atmosphere on Mars, including latitudinal effects and large bodies of water; and (8) the long-term stability of a thick atmosphere on Mars in the absence of tectonic recycling of sediments. Many of these questions can be studied with numerical models, but some require the further exploration of Mars.

If such investigations indicate that it is feasible to make Mars habitable, the motivation to do so may depend on the potential for life (human and non-human) on Mars, the economic payoff, and the nature of other large-scale efforts that provide habitats in space. We suggest that a key goal for future exploration of Mars should be to determine the feasibility of terraforming that planet. □

Christopher P. McKay is at the Space Science Division and Owen B. Toon at the Earth System Science Division, NASA Ames Research Center, Moffett Field, California 94035, USA. James F. Kasting is at the Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

1. Maddox, J. *Nature* **346**, 311 (1990).
2. Joos, F., Sarmiento, J. L. & Siegenthaler, U. *Nature* **349**, 772-775 (1991).
3. McKay, C. P. (ed.) *The Case for Mars II*, Am. Astronaut. Soc. Sci. Technol. Ser. Vol. 62 (Univelt, San Diego, 1985).
4. Stoker, C. R. (ed.) *The Case for Mars III*, Am. Astronaut. Soc. Sci. Technol. Ser. Vol. 74, 75 (Univelt, San Diego, 1985).
5. McKay, C. P. & Stoker, C. R. *Rev. Geophys.* **27**, 189-214 (1989).
6. Pollack, J. B., Kasting, J. F., Richardson, S. M. & Pollackoff, K. *Icarus* **71**, 203-224 (1987).
7. McKay, C. P. *Adv. Space Res.* **6**, 269-285 (1986).
8. Averner, M. M. & MacElroy, R. D. (eds) *On the Habitability of Mars*, NASA SP-414 (US Government Printing Office, Washington DC, 1976).
9. Burns, J. A. & Harwit, M. *Icarus* **19**, 126-130 (1973).
10. Sagan, C. *Icarus* **20**, 513-514 (1973).
11. Oberg, J. E. *New Earths* (Stackpole, Harrisburg, Pennsylvania, 1981).
12. McKay, C. P. *J. Brit. Interplanet. Soc.* **35**, 427-433 (1982).
13. Lovelock, J. E. & Allaby, M. *The Greening of Mars* (Warner, New York, 1984).
14. Fogg, M. J. *J. Brit. Interplanet. Soc.* **42**, 577-582 (1989).
15. Haynes, R. H. in *Moral Expertise* (ed. MacNiven, D.) 161-183 (Routledge, London, 1990).
16. McKay, C. P. in *Moral Expertise* (ed. MacNiven, D.) 184-197 (Routledge, London, 1990).
17. Dole, S. H. *Habitable Planets for Man*, 2nd edn (Elsevier, New York, 1970).
18. Bauer, J. A. & Martha, P. Z. *Pflanzenphysiol.* **103**, 445-450 (1981).
19. Lloyd, N. D. H., Canvin, D. T. & Culver, D. A. *Plant Physiol.* **59**, 936-940 (1977).
20. Salisbury, F. B. & Ross, C. W. *Plant Physiology* (Wadsworth, Belmont, California, 1985).
21. Lovelock, J. E. & Whitfield, M. *Nature* **296**, 561-563 (1982).
22. Rogers, H. H., Thomas, J. F. & Bingham, G. E. *Science* **220**, 428-429 (1983).
23. Hickleton, P. R. & Jolliffe, P. A. *Can. J. Bot.* **58**, 2181-2189 (1980).
24. Seckbach, J., Baker, F. A. & Shugarman, P. M. *Nature* **227**, 744-745 (1970).
25. Drew, M. C. *Curr. Adv. Plant Sci.* **36**, 1-14 (1979).
26. Klingler, J. M., Mancinelli, R. L. & White, M. R. *Adv. Space Res.* **9**, 173-176 (1989).
27. Billings, C. E. in *Bioastronautics Data Book* (eds Parker, J. F. Jr & West, V. R.) NASA SP-3006, 1-34; 35-64 (Washington DC, 1973).
28. Welch, B. E., Morgan, T. E. Jr & Clamann, H. G. *Federation Proc.* **22**, 1053-1056 (1963).
29. MacEwen, J. D. in *Bioastronautics Data Book* (eds Parker, J. F. Jr & West, V. R.) NASA SP-3006, 455-488 (Washington DC, 1973).
30. Lovelock, J. E. *Gaia* (Oxford University Press, Oxford, 1979).
31. Roth, E. M. in *Atmosphere in Space Cabins and Closed Environments* (ed. Kemmermeyer, K.), 13-31 (Merideth, New York, 1964).
32. Anders, E. & Grevesse, N. *Geochim. cosmochim. Acta* **53**, 197-214 (1989).
33. Owen, T. et al. *J. geophys. Res.* **82**, 4635-4639 (1977).
34. Pollack, J. B. *Icarus* **37**, 479-533 (1979).
35. Cess, R. D., Ramanathan, V. & Owen, T. *Icarus* **41**, 159-165 (1980).
36. Hoffert, M. I., Callegari, A. J., Hsieh, C. T. & Ziegler, W. *Icarus* **47**, 112-129 (1981).
37. Postawko, S. E. & Kuhn, W. R. *Proc. 16th Lunar Planet. Sci. Conf., J. geophys. Res.* **91**, D431-D438 (1986).
38. Kieffer, H. H. et al. *J. geophys. Res.* **82**, 4249-4292 (1977).
39. Kong, T. Y. & McElroy, M. B. *Icarus* **32**, 168-189 (1977).
40. Kasting, J. F. & Walker, J. C. G. *J. geophys. Res.* **86**, 1147-1158 (1981).
41. Kasting, J. F., Pollack, J. B. & Crisp, D. J. *atmos. Chem.* **1**, 403-428 (1984).
42. Oyama, V. I. & Berdahl, B. J. *J. geophys. Res.* **82**, 4669-4676 (1977).
43. Kasting, J. F. & Ackerman, T. P. *Science* **234**, 1383-1385 (1986).
44. Ramanathan, V., Cicerone, R. J., Singh, H. B. & Kiehl, J. T. *J. geophys. Res.* **90**, 5547-5566 (1985).
45. Weast, R. C. *CRC Handbook of Chemistry and Physics* (CRC, West Palm Beach, Florida, 1978).
46. Hammit, J. K. *Nature* **330**, 711-716 (1987).
47. Gierasch, P. J. & Toon, O. B. *J. Atmos. Sci.* **30**, 1502-1508 (1973).
48. Sagan, C., Toon, O. B. & Gierasch, P. J. *Science* **181**, 1045-1049 (1973).
49. Paige, D. A., Herkenhoff, K. E. & Murray, B. C. *J. geophys. Res.* **95**, 1319-1335 (1990).
50. Fanale, F. P. & Cannon, W. A. *J. geophys. Res.* **84**, 8404-8414 (1979); **79**, 3397-3402 (1974).
51. Fanale, F. P., Salvail, J. R., Banerdt, W. B. & Saunders, R. S. *Icarus* **50**, 381-407 (1982).
52. Zent, A. P. & Fanale, F. P. *J. geophys. Res.* **91**, D439-D445 (1986).
53. Toon, O. B., Pollack, J. B., Ward, W., Burns, J. A. & Bilski, K. *Icarus* **44**, 552-607 (1980).
54. Zent, A. P., Fanale, F. P. & Postawko, S. E. *Icarus* **71**, 241-249 (1987).
55. Valenzuela, D. P. & Meyers, A. L. *Adsorption Equilibrium Data Handbook* (Prentice Hall, Englewood Cliffs, New Jersey, 1989).
56. Lewis, J. S. *Earth planet. Sci. Lett.* **15**, 286-290 (1972).
57. Anders, E. & Owen, T. *Science* **198**, 453-465 (1977).
58. Rasool, S. I. & LeSergeant, L. *Nature* **266**, 822-823 (1977).
59. Clark, B. C. & Baird, A. K. *Geophys. Res. Lett.* **6**, 811-814 (1979).
60. McElroy, M. B., Kong, T. Y. & Yung, Y. L. *J. geophys. Res.* **82**, 4379-4388 (1977).
61. Pollack, J. B. & Black, D. C. *Science* **205**, 56-59 (1979); *Icarus* **51**, 169-198 (1982).
62. Dreibus, G. & Wänke, H. *Icarus* **71**, 225-240 (1987).
63. Becker, R. H. & Pepin, R. O. *Earth planet. Sci. Lett.* **69**, 225-242 (1984).
64. Pepin, R. O. *Rev. Geophys.* **25**, 293-296 (1987).
65. Bogard, D. D. & Johnson, P. *Science* **221**, 651-654 (1983).
66. Greeley, R. *Science* **236**, 1653-1654 (1987).
67. Carr, M. H. *Icarus* **68**, 187-216 (1986).
68. Liu, S. & Donahue, T. M. *Icarus* **28**, 231-246 (1976).
69. Fox, J. L. & DaGarno, A. *J. geophys. Res.* **88**, 9027-9032 (1983).
70. Owen, T., Maillard, J. P., de Bergh, C. & Lutz, B. L. *Science* **240**, 1767-1770 (1988).
71. Yung, Y. L. et al. *Icarus* **76**, 146-159 (1988).
72. Carr, M. H. *Icarus* **87**, 210-227 (1990).
73. Melosh, H. J. & Vickery, A. M. *Nature* **338**, 487-489 (1989).
74. Meyer, T. R. & McKay, C. P. *J. Brit. Interplanet. Soc.* **42**, 147-160 (1989).
75. Squyres, S. W. & Carr, M. H. *Science* **231**, 249-252 (1986).
76. Pollack, J. B. et al. *J. geophys. Res.* **95**, 14595-14628 (1990).
77. Toulmin, P. III, Baird, A. K., Clark, B. C., Keil, K. & Rose, H. J. *Jr Science* **194**, 81-84 (1976).
78. Toulmin, P. et al. *J. geophys. Res.* **82**, 4625-4634 (1977).
79. Prinn, R. G. & Fegley, B. Jr. *Ann. Rev. Earth planet. Sci.* **15**, 171-212 (1987).
80. Dreibus, G. & Wänke, H. *Meteoritics* **20**, 367-381 (1985).
81. Clark, B. in *Mars* (University of Arizona Press, Tucson, 1991).
82. Whittaker, R. H. *Communities and Ecosystems* 2nd edn (Macmillan, New York, 1975).
83. Turekian, K. K. & Clark, S. P. Jr. *J. Atmos. Sci.* **32**, 1257-1261 (1975).

Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA

B. F. Luisi^{*†}, W. X. Xu^{*}, Z. Otwinowski^{*}, L. P. Freedman^{†‡}, K. R. Yamamoto[‡] & P. B. Sigler^{*}

^{*} Department of Molecular Biophysics and Biochemistry, and the Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06511, USA

[‡] Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143-0448, USA

Two crystal structures of the glucocorticoid receptor DNA-binding domain complexed with DNA are reported. The domain has a globular fold which contains two Zn-nucleated substructures of distinct conformation and function. When it binds DNA, the domain dimerizes, placing the subunits in adjacent major grooves. In one complex, the DNA has the symmetrical consensus target sequence; in the second, the central spacing between the target's half-sites is larger by one base pair. This results in one subunit interacting specifically with the consensus target half-site and the other nonspecifically with a noncognate element. The DNA-induced dimer fixes the separation of the subunits' recognition surfaces so that the spacing between the half-sites becomes a critical feature of the target sequence's identity.

ON binding hormone, the glucocorticoid receptor translocates from the cytoplasm to the nucleus where it binds to specific DNA sequences (glucocorticoid response elements, GREs) and modulates transcription rates of target genes^{1,2}. The glucocorticoid receptor is a member of the nuclear receptor superfamily, which includes the receptors for the steroid hormones, retinoids, peroxisomal activators³, vitamin D, thyroid hormones^{4,5}, certain regulators of early development in *Drosophila*^{6,7}, and other proteins of unknown function^{5,8}. Members of this superfamily have a common amino-acid sequence organization composed of discrete functional domains for ligand binding, DNA binding and transcriptional regulation. The DNA-binding domain is sufficient for selective interaction with specific response elements⁹⁻¹².

The DNA-binding domain of the nuclear receptors is characterized by a pattern of eight cysteines which, for the glucocorticoid receptor, coordinate two Zn²⁺ ions, each with tetrahedral geometry¹³. In coordinating the metal, peptide loops are formed that are superficially similar to those of the 'Zn-fingers' described originally for the transcription factor IIIA (TFIIIA) from *Xenopus*^{14,15} and since inferred to exist in a wide variety of nucleic acid-binding proteins^{16,17}. The substructures nucleated by Zn coordination in the nuclear receptors are quite different from TFIIIA-type Zn-fingers: in the TFIIIA fingers, the Zn is coordinated by two cysteines and two histidines, rather than by four cysteines. Nuclear magnetic resonance spectroscopy has indicated that the Zn-fingers of the receptor and TFIIIA type fingers differ structurally¹⁸⁻²¹, which is borne out in this study. TFIIIA fingers act as independent, conformationally stable

structural units, each contributing to DNA binding²², whereas the Zn-fingers of the nuclear receptors fold together as part of a larger, unified globular domain.

Here we characterize complexes of the glucocorticoid receptor's DNA-binding domain with two oligonucleotides, which differ only in their half-site spacing. The principal study at 2.9 Å employs a DNA target in which the identical half-sites of an idealized glucocorticoid response element²³ are symmetrically disposed around a spacer of four base pairs (GRE_{4S}) instead of the naturally occurring three (Fig. 1b). This was done to provide a target with exact symmetry using a single oligonucleotide, an experimental expedient that seemed to be justified on the basis of evidence that the protein was monomeric in solution²⁴ and bound noncooperatively to symmetrical DNA targets (L.P.F., unpublished results). The crystal structure gave a surprising and mechanistically important result: namely that, on binding to DNA, the domains dimerized strongly with the subunits spaced so that they would interact correctly with half-sites separated by the normal three base pairs. With a spacing of four bases, only one subunit seemed to interact properly with one half of the target. The extra base pair in the centre displaced the second half-site by one base pair, causing the second subunit to interact with a noncognate sequence. We propose that one domain forms a 'specific' interaction and the second a 'non-specific' one.

Encouraged by our result with GRE_{4S} and other biochemical results indicating that the DNA-binding domain of both the glucocorticoid and oestrogen receptor binds cooperatively to pseudosymmetric response elements with the natural three base-pair spacers^{25,26}, we crystallized the complex containing the pseudosymmetric target, GRE_{3S}. The design principles for crystallizing the GRE_{3S} complex were provided by an analysis of the crystal packing contacts of the GRE_{4S} complex. A preliminary analysis (4.0 Å resolution) of the complex with a three base-pair spacing indicates that both subunits correctly face the half-sites. The two crystal structures are compared and a model is proposed to explain the role of DNA-induced dimerization in producing the graded specific affinity of the glucocorticoid receptor for different DNA targets.

Crystallization, structure and refinement

An 86-amino-acid fragment of the rat glucocorticoid receptor encompassing amino-acid residues 440 to 525 of the DNA-binding region was expressed in *Escherichia coli* (Fig. 1a; known hereafter as 440-525). Fragment 440-525 was co-crystallized with the self-complementary oligonucleotide shown in Fig. 1b. This sequence differs from the consensus palindromic GRE only by the insertion of a single base between the two half-sites to produce a spacer of four rather than three, resulting in a symmetrical duplex (GRE_{4S}).

The crystal structure was solved by multiple isomorphous replacement (MIR). The heavy-atom derivatives each have bromo-uracil or iodo-uracil in substitution for thymine at a unique position of the GRE_{4S} sequence (Table 1). The asymmetric unit of the 440-525/GRE_{4S} crystal consists of one DNA duplex and a dimer of fragment 440-525. The positions of the four Zn ions of the asymmetric unit, which correspond to the

[†] Present addresses: MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JT, UK (B.F.L.); Cell Biology and Genetics Program, Sloan-Kettering Institute, 1275 York Avenue, New York 10021, USA (L.P.F.).

highest electron density in the protein, were confirmed by their anomalous scattering.

The model was built into a map that had been improved by solvent flattening^{27,28}. It was refined by conventional least-squares refinement using the programs NUCLSQ and PROFFT (refs 29–32) and, in parallel, by simulated annealing, using the program X-PLOR (ref. 33). It was clear from the early maps that the DNA structure is B-type and that each Zn atom is tetrahedrally coordinated. Owing to the low resolution, the pucker of the furanoside rings were restrained to the C2'-endo conformation generally characteristic of the B family, and restraints were also applied to maintain the Zn–S centres to have tetrahedral symmetry and the bond distances that were found by spectroscopy¹³.

The sequences of GRE_{S4} and the dimer both have twofold symmetry, but this symmetry was not imposed in the refinement. It was subsequently found that the dyad of the dimeric protein and of the GRE_{S4} duplex do not coincide, as discussed later. The agreement between the calculated and observed X-ray structure factors of the final model is 19.6%. The root-mean-square fit of the peptide backbone of the two independently refined subunits is 0.6 Å. A synopsis of the stereochemistry of the final model is given in Table 2, and the quality of the electron density map is shown in Fig. 2.

Using the same conditions as those described for preparing the 440–525/GRE_{S4} crystals, we have also obtained crystals of a complex of 440–525 with a pseudo-palindromic GRE (referred to as GRE_{S3}) in which the spacing between half-sites is three bases, as in the consensus sequence (Fig. 1b). So far, the GRE_{S3}

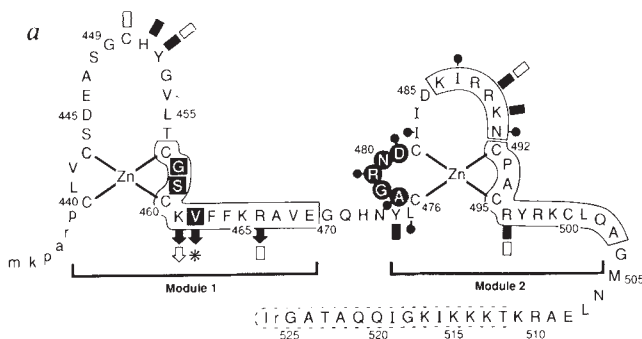


FIG. 1 Sequences of molecules used for crystallographic analyses. a, Fragment 440–525, showing the Cys–Zn connectivity and α -helical segments, which are boxed. The numbering convention corresponds to that of the full-length native receptor. The relationship of structure and function is best described in terms of substructures called 'modules' which are indicated by the brackets. Residues that make dimer interface contacts are indicated by the solid dot (475, 477, 479, 481, 483, 487 and 491). Residues making phosphate contacts at the specific (451, 452, 474, 490, 489 and 496) and the nonspecific sites (450, 452, 466, 489 and 496) are indicated by solid and open rectangles, respectively. The solid and open arrows indicate base contacts at the specific and nonspecific site, respectively. The asterisk indicates that the contact between Val 462 and the base is not made at the nonspecific site. A disordered section at the carboxy-terminus is indicated by the dashed lines. Three amino acids that direct the discrimination of glucocorticoid and oestrogen response elements are indicated by white lettering in the solid boxes (458, 459 and 462) and those that discriminate between oestrogen and thyroid response elements are in discs (478–481; refs 9–12). Cloning artefacts from the expression vector construct are indicated by lower-case letters. Substitution of Pro for the naturally occurring Leu at position 439 does not affect activity *in vivo* or *in vitro*³⁷. b, The sequences of the DNA used for crystallizing the half specific/half nonspecific complex (GRE_{S4}), with four bases (within brackets) separating the hexameric half-sites (underlined), and the fully specific complex (GRE_{S3}), in which three bases separate the elements. The numbering convention employed here to refer to bases and phosphates uses the dyad as origin, as indicated. For the GRE, ERE (ref. 23) and thyroid hormone response element (TRE) consensus⁴³, the hexameric half-sites are indicated by arrows. The vertical arrow indicates the dyad (for GRE_{S4} and TRE) or pseudo-dyad (for the other sequences).

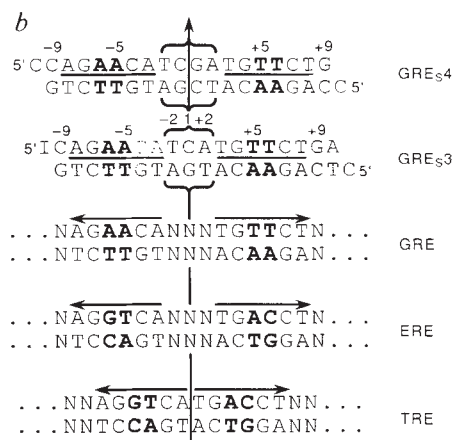
complex crystals are smaller and less well-ordered than the GRE_{S4} crystals, but a preliminary model was obtained for the 440–525/GRE_{S3} complex with an incomplete 4.0 Å data set using a rigid body molecular replacement search (X-PLOR; see legend to Table 1). The independent rigid bodies used were the two subunits and two GRE half-sites from the refined 2.9 Å GRE_{S4} model.

Crystal packing

Adjacent GRE_{S4} molecules are linked end-to-end through a triple helix in which the 5'-overhanging C (protonated) of one duplex (Fig. 1b) makes a Hoogsteen-like interaction³⁴ with the G of the 3'-terminal base pair of the neighbouring duplex. The triple helix was exploited in designing oligonucleotides used to obtain 440–525/GRE_{S3} crystals, which are almost isomorphous with those of the GRE_{S4} complex. As planned, the different 5' ends of GRE_{S3} aligned the complexes directionally (which are asymmetric with respect to the duplex) by forming distinctive I: A and C⁺: G Hoogsteen interactions. Triple-helix interactions have also been observed in the crystal structures of the *met* repressor-operator³⁵ and the catabolic activator protein-operator complexes³⁶. The triple helix may prove an important consideration when designing sequences for crystallization.

Tertiary structure of the protein

The principal secondary structural elements of 440–525 are illustrated schematically in Fig. 3a. The two independently refined 440–525 molecules of the crystallographic asymmetric unit superimpose within the error of the models. The protein



METHODS. Fragment 440–525 was purified by a modification of the method used to isolate a construction of relative molecular mass 19,000 of the DNA-binding domain¹³. A cell lysate was fractionated with ammonium sulphate and the protein was enriched by elution from CM-Sepharose CL6B (Pharmacia) with a linear gradient from 0.05–0.6 M NaCl in 50 mM Tris-Cl, pH 7.5, followed by elution from a mono-S FPLC column (Pharmacia) with a 0.05–0.46 M NaCl gradient in 20 mM sodium phosphate, pH 7.5. The amino acids of the amino and carboxy termini were confirmed by sequencing. The extinction coefficient of fragment 440–525 was found by amino-acid analysis: 205 μ M protein has an absorbance of 1.0 in a 1.0-cm light path at 280 nm. Oligonucleotides were synthesized by the phosphoramidite method and purified by reverse-phase HPLC on a C4 resin (Vydak), followed by FPLC at pH 11.8 with a mono-Q matrix (Pharmacia). DNA concentration was calculated using extinction coefficients for individual nucleosides and compensating for hypochromic extinction. Crystals of the 440–525/GRE_{S4} complex, its heavy-atom derivatives, and of the 440–525/GRE_{S3} complex were grown by the method of vapour diffusion from hanging droplets. The optimal conditions are as follows: a droplet containing 0.07 mM DNA duplex, 0.14 mM 440–525, 1.8 mM MgSO₄, 25 mM NaCl, 18 mM sodium cacodylate, pH 6.0, 1.8 mM spermine and 2 μ M ZnCl₂, was equilibrated against a reservoir containing 0.2 M NaCl, 24 mM sodium cacodylate, pH 6.0, and 8% (by volume) methyl pentane diol or ethanol at 8 °C. Crystals grew within a week to their maximal extent, typically with dimensions 0.2 $\times$ 0.3 $\times$ 0.5 mm.

FIG. 2 Experimental X-ray data. Stereoscopic view of the electron density map in the vicinity of the tip of the amino-terminal finger, showing its interaction with the phosphate backbone. His 451 and the main-chain amide of Tyr 452 donate to the phosphate oxygens of A(-8) and the hydroxyl group of Tyr 452 donates to the phosphate oxygen of G(-7). The map was calculated using minimum bias coefficients⁵³, and the phases combined from calculated and MIR values by Sim weighting.

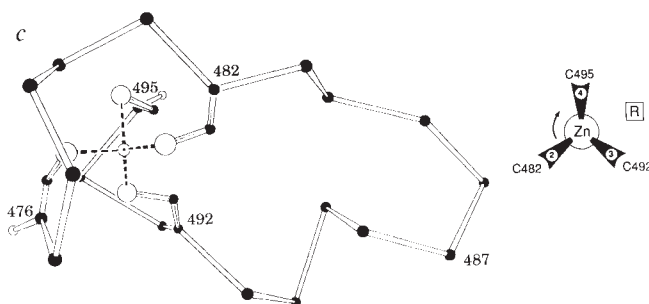
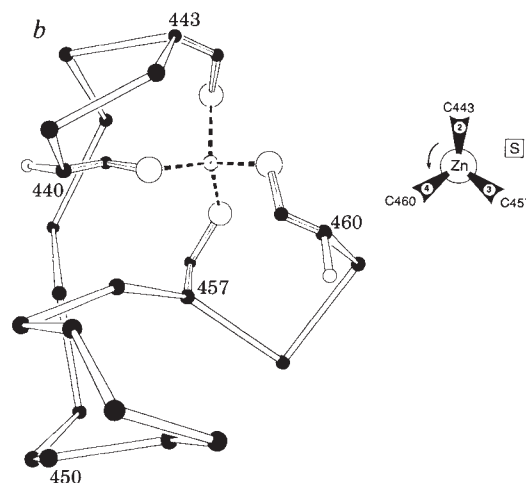
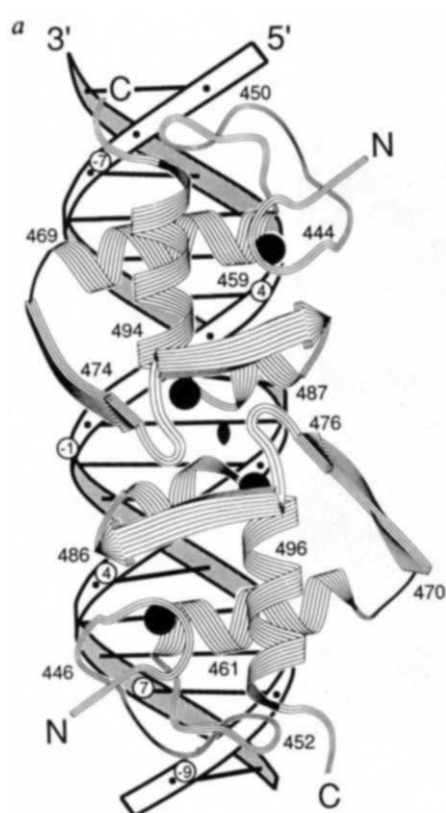
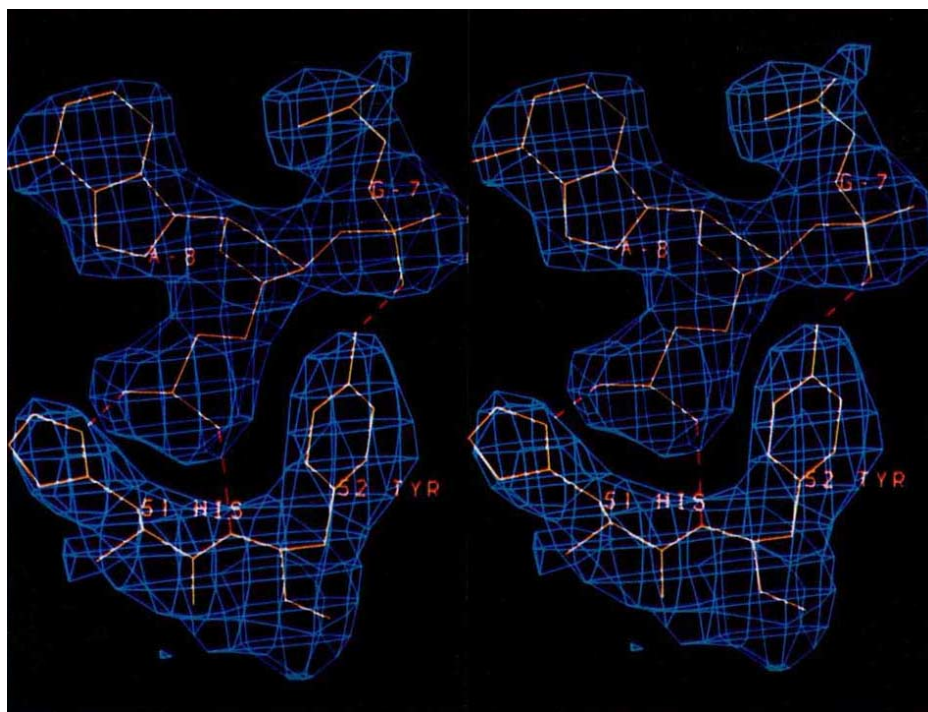


FIG. 3 Schematic representation of the protein and a comparison of Zn-fingers. *a*, A ribbon representation⁵⁴ of the glucocorticoid receptor/GRE_{S4} complex, looking approximately down the protein's dyad toward the interactive surface of the DNA. The amino (N) and carboxy (C) termini are indicated, and the Zn ions are represented by discs. *b*, The α -carbon trace of the amino-terminal finger (left), and the configuration of the Zn-S geometry (right). The view is rotated 180 degrees relative to *a*. The configuration is *S*, using the sequence-based convention defined by Berg (ref. 16). As in

other metal binding proteins, there are hydrogen-bonding interactions between the coordinating cysteines and the peptide backbone. There is a short segment of anti-parallel β structure in the tip of the finger. *c*, The carboxy-terminal finger (left) and Zn-S configuration (right), which is *R*—the opposite of that of the amino-terminal finger. The finger is viewed in the same orientation as in *a*. The α helix within the finger is clear in the map, and refines with definite but irregular helical stereochemistry.

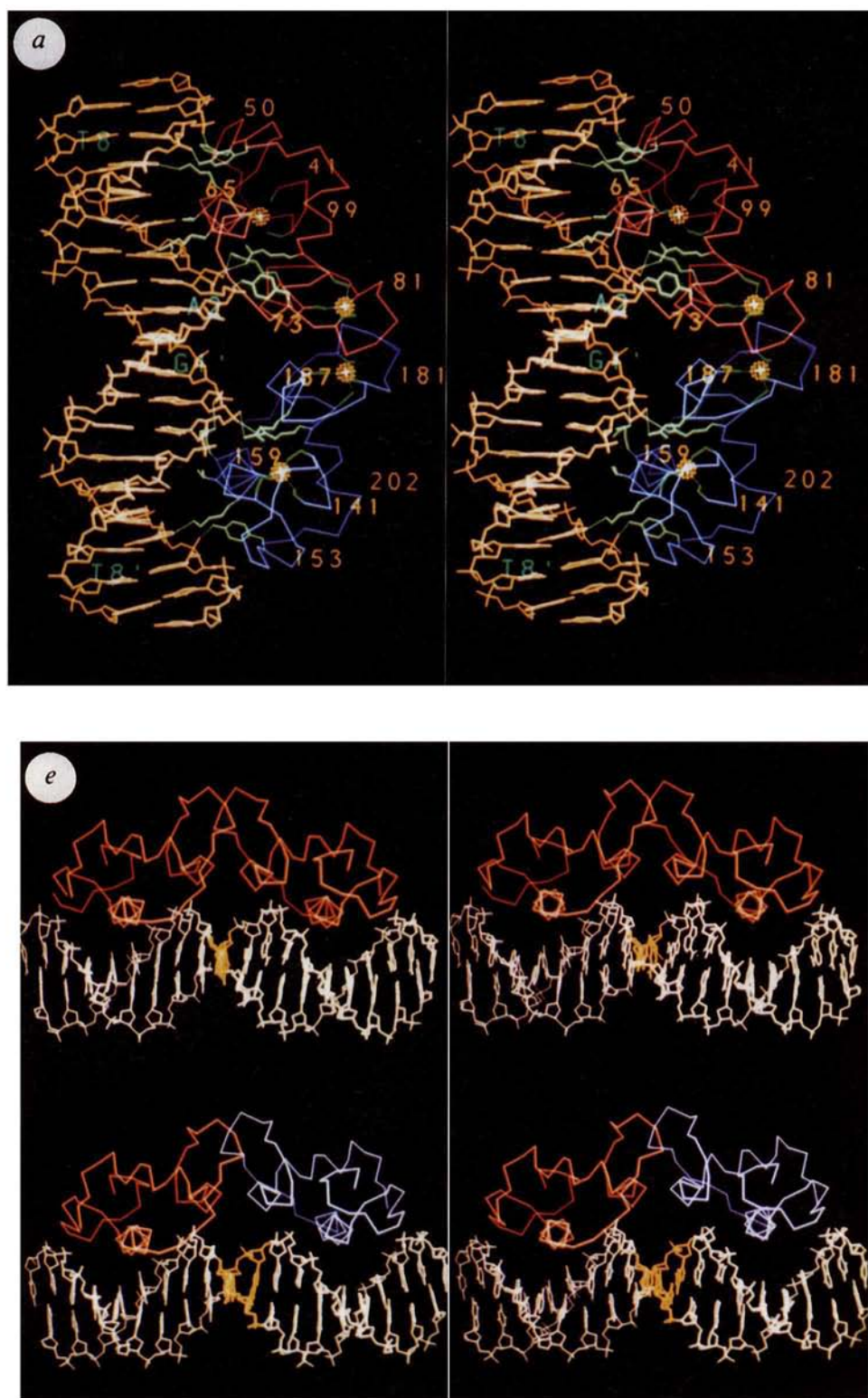
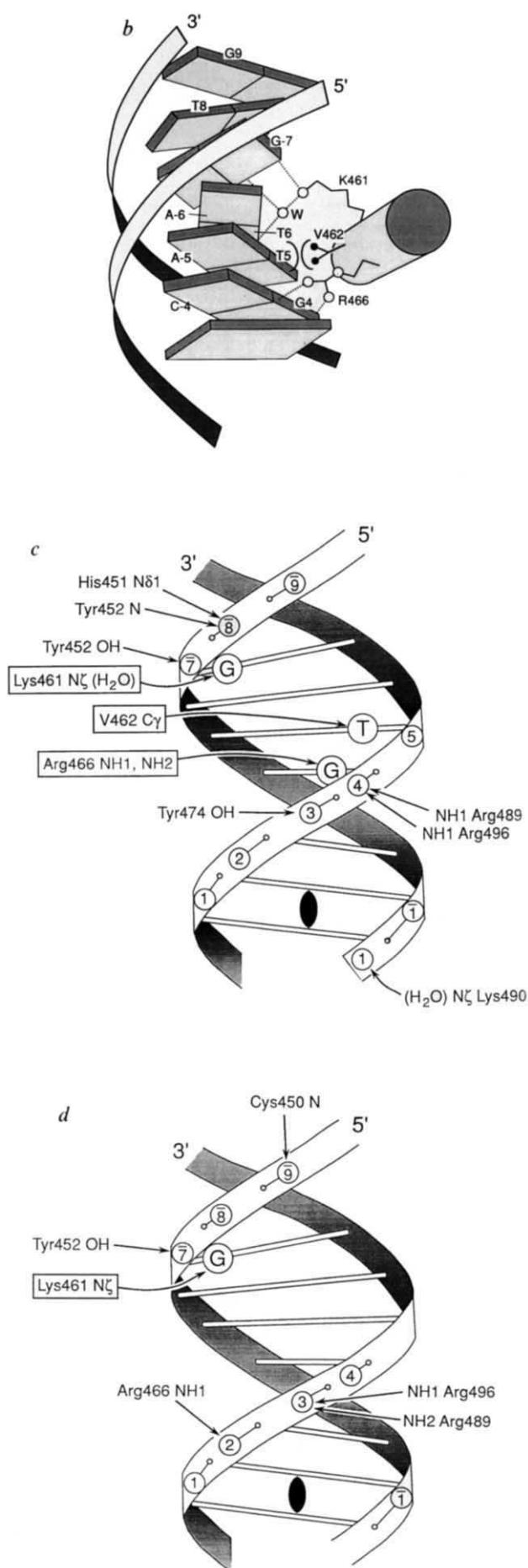


FIG. 4 The 440-525/DNA complex. *a*, Stereoscopic view of the GRE_{S4} complex. Only the α -carbon tracing of the protein is shown for clarity, except for the side chains interacting with the bases and phosphates, and the cysteine residues coordinating the Zn ions, which are shown in green. The subunit making specific contacts is shown in red, and the nonspecific subunit is in blue. Zn ions are indicated by the spheres. Dashed lines indicate direct hydrogen bonds between the protein and DNA. *b*, Perspective drawing of the base contacts in the major groove of the specific half-site. *c*, Schematic diagram of the GRE_{S4} specific and *d*, nonspecific sites, identifying the phosphate and base contacts. The dyad of the sequence is indicated by the solid oval. *e*, A stereopair comparing the 440-525/GRE_{S3} (top) and 440-525/GRE_{S4} (bottom) complexes. The nonspecific subunit of the GRE_{S4} complex is shown in blue and the central base pairs between which the dyad of GRE_{S4} passes are indicated in yellow. For the GRE_{S3} complex, the base pair through which the dimeric protein's dyad passes is shown in yellow. The approximate dyad axis of the GRE_{S4} passes between the base pairs at positions -1 and 1, whereas the axis of the dimer is displaced axially by roughly one-half of a base-pair step towards the subunit that makes contacts with the DNA bases and azimuthally by 12°. This is consistent with our proposal that (1) one subunit interacts specifically with a half-site of GRE_{S4} and (2) the distance between protein binding sites is maintained at 3 bases, as found in the consensus GRE (Fig. 1*b*), and cannot be relaxed to four bases to accommodate the two half-sites of GRE_{S4}. There are two possible orientations of the 440-525/GRE_{S4} and GRE_{S3} complexes, but the crystal lattice (protein-protein) contacts favour one.

assumes a compact, globular fold and can be divided into two substructures or 'modules', each being nucleated by a Zn coordination centre and followed by an amphipathic α helix (Fig. 1*a*). Interestingly, these two modules are encoded by separate exons in the gene for the glucocorticoid receptor (M. Jacobson and K.R.Y., unpublished results), as well in the genes of several other members of the nuclear receptor superfamily⁵.

The two substructures are joined together mainly through the interaction of the aromatic side chains of their amphipathic helices (Phe 463, Phe 464 and Tyr 497) which, together with

Tyr 452 from the large loop of the first finger, and Tyr 474 from the β strand linking the two modules, form an aromatic cluster. A shell of hydrophobic residues (Cys 500, Met 505, Pro 493 and the hydrocarbon side chain of Arg 496) surround this aromatic cluster. The resulting hydrophobic core stabilizes the globular fold and fixes the relative orientation of the two substructures. Except for Pro 493 and Tyr 497, the residues constituting the core are highly conserved in the superfamily. Non-conservative substitutions at several of these positions of 440-525 result in loss of function³⁷.



In agreement with spectroscopic¹³ and site-directed mutagenesis results³⁸, the crystal structure of the complex shows that the protein coordinates two zinc ions, each by a tetrahedral configuration of sulphur atoms from four cysteine residues (identified in Fig. 1a). As is found in *Xenopus* Xfin (ref. 19) and in other structurally characterized metal-binding proteins¹⁶, the chiral Zn-sulphur centre of the amino-terminal finger assumes the *S* configuration (Fig. 3b), presumably as a result of conformational restrictions of the peptide backbone. Surprisingly, the Zn coordination geometry of the carboxyl-terminal finger assumes the *R* configuration, or opposite hand (Fig. 3c). This may reflect the greater conformational freedom afforded by the relatively longer peptide linking the first and second coordinating residues (Cys 476 and Cys 482; see Fig. 1a).

The two subdomains of 440–525 differ structurally and functionally. Where the main loop of the amino-terminal finger has a short segment of antiparallel β structure, the carboxy-terminal finger has a distorted α helix. The β sheet of the amino-terminal finger helps to orient the residues that make phosphate contacts (Fig. 2), and its second pair of Zn-coordinating cysteines begin an α helix that provides all of the base contacts in the major groove of the GRE. The carboxy-terminal subdomain provides phosphate contacts, the entire dimerization interface (Figs 3a, 6), as well as a site implicated in positive control of transcription³⁷. Most of the intersubunit contacts are made by residues in the loop between Cys 476 and Cys 482, a region referred to as the 'D box' for its postulated role in dimerization¹¹ (Figs 1a and 6). The conformation of this loop is a reverse β turn and is maintained by the *R*-configuration of the Zn centre. The role of the carboxy-terminal substructure in providing all of the known protein-protein contacts may be the underlying functional reason for the distinctive chirality of the metal centre. There is a β strand in the main loop of the carboxy-terminal finger which exposes the hydrophobic residues Ile 483 and Ile 487 to the solvent, a situation which is sufficiently unusual in protein structures to suggest a potential role for these residues in contacting other domains of the receptor or even other proteins³⁷.

The secondary and tertiary structures of both Zn-nucleated substructures in 440–525 differ significantly from those of the TFIIIA-type fingers of *Xenopus* Xfin (ref. 19), yeast ADR1 (ref. 18), and Zif268 (ref. 22) and also from the Zn-binding module of the gag protein p55 from human immunodeficiency virus³⁹. Unlike the Zn-nucleated modules of 440–525, the TFIIIA-type Zn fingers are stabilized individually by an extensive hydrophobic core. Moreover, they assume a consistent conformation seemingly independent of their molecular context and the presence or absence of DNA²².

The NMR structures of 440–525 (ref. 20) and of the 11K DNA-binding domain from the oestrogen receptor²¹ have been determined in the absence of DNA. There is no evidence of dimer formation even at the millimolar concentration of subunits used in the NMR experiment. The crystal and NMR studies show approximately the same tertiary structure, especially in the amino terminal module where the root-mean-square fit of the first module (α carbon residues 440–469) in the crystal structure to that in the NMR structures of GR440–525 is 2.0 Å, compared with 1.55 Å for the fit of the independent energy minimized solutions for the NMR structure²⁰. But, neither of the NMR structures reveals the hydrogen-bonding interactions of the short segment of antiparallel β structure in the amino terminal finger, or of the distorted α helix in the second finger, both of which are observed in the X-ray structure of the dimeric complex with DNA. As these two segments make DNA and/or dimer contacts in the crystal structures, the differences between the NMR and crystal structures may arise from a stabilization of secondary structure upon formation of the complex.

There is a much poorer correspondence in structure between the second module in the crystal structure and its counterparts in the NMR structures of the glucocorticoid and oestrogen

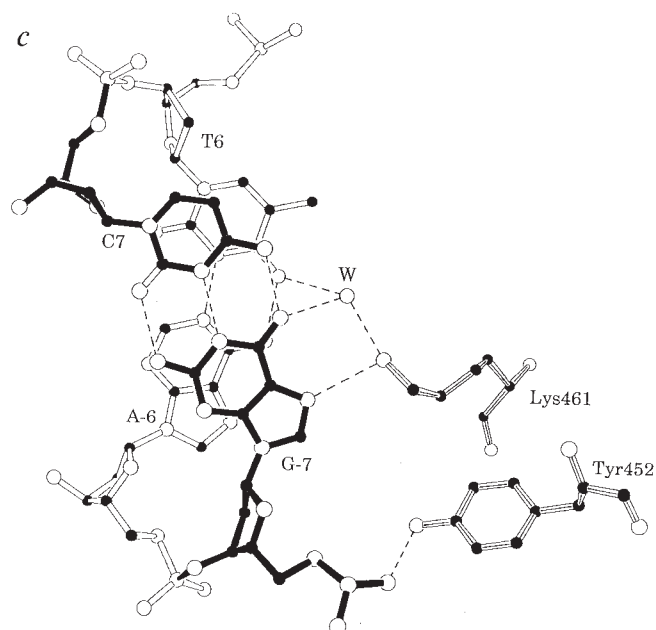
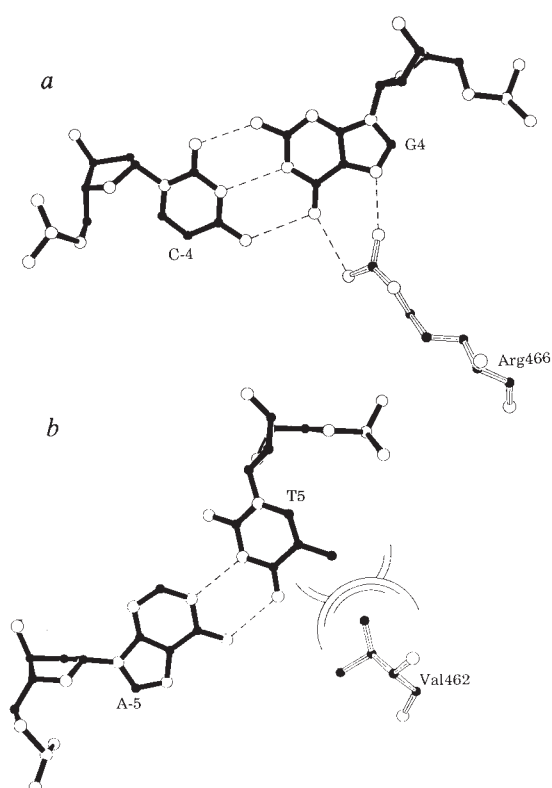


FIG. 5 The base-contacts of the 440–525/GRE_{S4} specific site. *a*, Arg 466 donates two hydrogen bonds to G(4). *b*, The methyl group of Val 462 makes a van der Waals contact with the 5-methyl group of T(5). *c*, Lys 461 donates a hydrogen bond to N7 and, through a water molecule, to O6 of G(–7) and O4 of T(6).

receptors. This module provides both phosphate contacts and the entire dimer interface. It is important to note in the descriptions that follow that the phosphate contacts of Tyr 474, Arg 489, Lys 490 and Arg 496 arise from the same substructure that forms the dimer interface. We suggest that these interactions help stabilize the conformation and orientation of the dimer interface, thereby causing cooperative enhancement of dimerization and DNA binding. Finally, the residues of the carboxy-terminus, 512 to 525, are poorly ordered in the X-ray structure and in the two NMR structures. This region, which is rich in basic amino

acids (Fig. 1*a*) is partially responsible for the nuclear localization of the receptor⁴⁰.

Protein–DNA interactions

Two molecules of fragment 440–525 bind to one face of the 18-base-pair DNA as a twofold rotationally symmetric dimer (Figs 3*a* and 4*a*). The interactive surface of each subunit engages successive major grooves. There are no protein contacts to the bases in the intervening minor groove. This mode of interaction is consistent with the methylation protection patterns of GREs

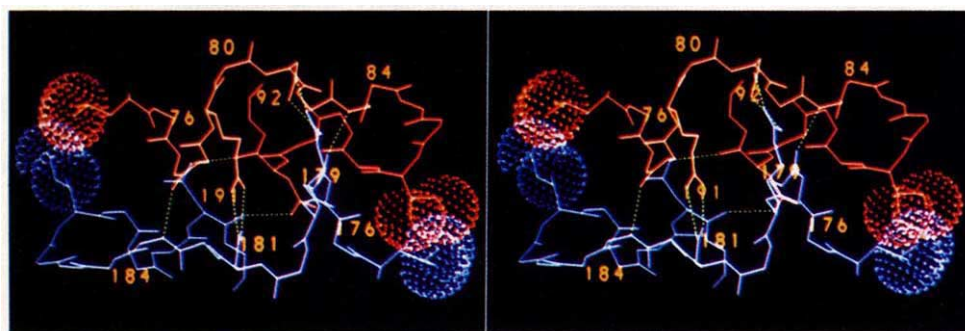
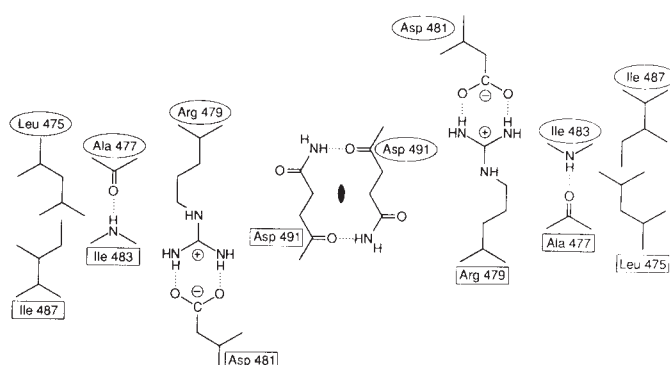


FIG. 6 The dimer interface. A stereopair showing the dimer interface viewed down the dyad axis. The DNA helical axis is vertical and the DNA-binding surface is on the 'far side' as in the ribbon diagram (Fig. 3*a*). The specifically bound subunit is shown in red. The residue numbers of the red subunit are 400 less than their true value, the numbers of the blue subunit are 300 less. A schematic diagram below shows the progression of dimer-stabilizing interactions in the stereopair with oval labels indicating the red subunit, rectangular labels the blue.



provided by the DNA-binding domain²⁵ and by the native receptor^{41,42} and agrees generally with the molecular arrangements modelled by Hard *et al.* (ref. 20) and Schwabe *et al.* (ref. 21) on the basis of NMR studies of the uncomplexed monomeric DNA-binding domains of the glucocorticoid and oestrogen receptors, respectively.

Contacts are made to the phosphate strand furthest from the dyad (Fig. 4a) by the large loop of the amino-terminal finger and to the strand closest to the dyad (Fig. 4a) by the carboxy-terminal module. At both specific and nonspecific half-sites, the amino-terminal α helix lies in the major groove and its side chains form the only contacts to the bases (Fig. 4a, b). Positioned over (but not contacting) the minor groove are the carboxy-terminal modules of the two subunits, which interact extensively with each other. The mode of interaction of 440–525 with DNA differs from that of the TFIIIA class of fingers, where each of a linked series of fingers forms an independent structure and the series spirals along the major groove directionally (that is, with no rotational symmetry), making discriminating interactions with a three-base DNA segment²².

Because the four-base-pair spacing between half-sites is one base pair greater than that of the consensus GRE, and because the dimer interface is tight, only one subunit of the 440–525/GRE_{S4} complex faces the target half-site sequence in GRE_{S4}; the second subunit is forced out of register with the target sequence by one base pair (Fig. 4a, e). The latter subunit, therefore, faces a sequence with no resemblance to the consensus hexamer. An analysis of the 4.0-Å structure of 440–525/GRE_{S3} shows this reasoning to be correct: within the resolution limits of the data, the GRE_{S3} complex is a twofold rotationally symmetric version of the specific interaction seen in the GRE_{S4} complex (Fig. 4e), and the dimer interface remains the same.

The resolution of the fully specific 440–525/GRE_{S3} is not adequate to examine the DNA/side chain interactions. We therefore turn to the specific site of the 440–525/GRE_{S4} complex, which is at sufficient resolution, to analyse the origin of sequence specificity. In the convention given in Fig. 1b, G is found at position 4 in most steroid response elements, including the consensus GRE. This base makes two hydrogen bonds with the conserved Arg 466 at the specific site (Fig. 5a), but not at the nonspecific one, where Arg 466 makes a phosphate contact (Fig. 4d). If Arg 466 is replaced by Lys or Gly, the protein no longer functions *in vivo* and binds GRE weakly *in vitro*^{37,38}. T at position 4 is a unique and consistent feature of GREs, and the methyl group of T(5) makes a favourable van der Waals contact with the side chain of Val 462 at the specific site (closest point 3.1 Å; Fig. 5b). At the nonspecific site, the Val is too far from the T (Fig. 4a). G(–7) is also found in most hormone-response elements, and it accepts one direct and one water-mediated hydrogen bond from an invariant Lys 461 in the specific site. This water molecule also contributes a hydrogen bond to O4 of T6, another base that specifies GR affinity (Fig. 5c). At the nonspecific site, Lys 461 can extend only far enough to donate one hydrogen bond to the corresponding G (Fig. 4d). Arg 466 and Lys 461 are absolutely conserved in the superfamily and, not surprisingly, their targeted bases, G(4) and G(–7), occur consistently in all four known classes of the hormone-response elements (reviewed in ref. 43).

A complete account of the phosphate contacts is given in Fig. 4c, d. Like most other DNA-binding proteins, the recognition surface of 440–525 is lashed down into the major groove by an array of phosphate interactions. Many of the same protein-phosphate contacts occur in the specific and nonspecific sites, although some contacts seem to be designed for specific interaction, such as the hydrogen bond between the N δ of His 451 and the phosphate of A(–8) (Fig. 2). The orientation and hydrogen-bonding polarity of His 451 are defined by a hydrogen-bonded network in the tip of the first finger, in which the N ϵ of His 451 accepts a hydrogen bond from the hydroxyl group of Ser 448, which in turn is an acceptor for amide 458. In the receptor

family sequences, His is conserved at position 451, and Ser or Thr always occur at position 448, suggesting the importance of this stabilizing network. At both specific and nonspecific sites, Arg 489 of the carboxy-terminal finger donates hydrogen bonds to the phosphate backbone and is buttressed in this function by hydrogen bonds to Asp 445 in the main loop of the amino-terminal finger, to the hydroxyl of Ser 459 (a residue that helps the glucocorticoid receptor differentiate GRE from an oestrogen-response element, ERE) and to the backbone carbonyl of Asp 480 of the D box (which contributes to dimerization). This four-way interaction helps unify the entire domain both structurally and functionally. Schena *et al.* (ref. 37) found that the substitution of Arg 489 with Lys results in decreased affinity of the receptor for GRE sequences *in vitro*. Although positively charged, the Lys side chain provides neither the same length nor the hydrogen-bonding pattern to make these multiple interactions. Arg 489 is conserved in the superfamily and its functional importance has been implicated by the observation that Arg is replaced by Gln in the human vitamin D receptor of patients with hereditary vitamin D-resistant rickets⁴⁴.

In the complex, GRE_{S4} has a B-type structure with no significant bend or distortion, and the overall helical repeat is 10.6 base pairs, which is close to that of random sequence DNA in solution⁴⁵. However, the major groove of the specific site is widened by 2 Å relative to the nonspecific site, which permits the amino-terminal module to be buried more effectively. Many of the other helical parameters of the two half-sites also differ significantly. This suggests that the target sequence not only must be in register axially with the dimer's determinants of specificity, but must also have the correct azimuthal orientation to permit the appropriate structural accommodation. The solvent excluded surface area created by forming the complex is 1,620 Å² at the specific site, which reflects a significantly more stable interaction than the 1,140 Å² excluded at the nonspecific site.

Quaternary structure of the dimer

Fragment 440–525 is monomeric in solution²⁴, and the structures of the GRE_{S4} and GRE_{S3} complexes show that 440–525 dimerizes on DNA binding through the interaction of its carboxy-terminal modules (Figs 3a, 4a and 6). As inferred from mutagenesis studies^{11,46}, the residues between Cys 476 and Cys 482—the D box—form important contacts between subunits (Fig. 1a). These include (1) a symmetrical pair of inter subunit salt bridges between Arg 479 and Asp 481 (the mutation of Arg 479 to Lys would conserve these salt bridges, and indeed this mutant has wild-type activity *in vivo*³⁷), and (2) a symmetrical pair of intersubunit backbone hydrogen bonds between carbonyl 477 and NH 483. Outside this box, a symmetrical pair of hydrogen bonds occurs between carbonyl of Asn 491 and the corresponding Asn 491 side chain (Fig. 6). The dimer interface is further stabilized by a hydrophobic contact between Ile 487 and Leu 475.

The rigid body refinement of the GRE_{S3} complex takes no account of the chemistry, yet places the individual subunits so as to preserve the same dimer interface as seen in the GRE_{S4} complex. The interface must therefore be sufficiently tight to lock the two subunits into a fixed relative orientation, despite the different manner in which they contact the GRE_{S3} and GRE_{S4} sequences (Fig. 4e). Although the DNA-binding domain is monomeric in solution, its dimerization is favoured upon GRE binding for two reasons. First, the subunits bind contiguously and with appropriately restricted orientations. Thus, the entropy penalty of association is reduced by binding to DNA and especially to sequences with naturally spaced half-sites. Second, contacts between residues in the carboxy-terminal substructure and the DNA backbone appear to stabilize the conformation that supports dimerization. Therefore, tight DNA-induced dimerization explains the observation that 440–525 binds cooperatively to the half-sites of consensus GREs, as in GRE_{S3}

TABLE 1 Quality of MIR phases

Resolution (Å): Figure of merit*:	9.2	6.9	5.5	4.5	3.9	3.3	3.0	
	0.80	0.84	0.78	0.66	0.49	0.35	0.20	
Phasing power† of each derivative‡								Unique reflections
−2 (Br-U)	1.38	1.61	1.38	1.07	0.73	0.49	0.36	8,123
−2 (I-U)	—	1.52	1.45	1.09	0.79	0.77	0.66	5,620
+3 (Br-U)	0.92	1.15	1.09	0.83	0.54	0.34	0.23	8,175
+6 (Br-U)	0.84	1.04	0.90	0.63	0.42	0.30	0.20	7,582
+8 (I-U)§	1.43	1.25	1.25	1.00	0.58	0.37	0.21	7,419

In preparation for data collection, crystals were transferred through a series of buffered solutions (30 mM sodium cacodylate, pH 6.0, 1.5 mM MgSO₄, 2.0 μM ZnCl₂) of increasing methyl pentane diol concentration, finishing at 30% by volume of MPD. Crystal slippage and mosaicity were serious problems for data collection. Slippage was overcome by including 0.1% by weight of agarose (Sigma type VII) in the final stabilizer so that an immobilizing gel would develop when the temperature was brought below 4 °C. Crystals were mounted in tapered capillaries such that they were wedged by the glass wall and immersed in the agarose stabilizing solution. As they were continuously exposed to stabilizer, the crystals suffered less mosaic spread. A similar method of mounting crystals was used in studies of nucleosome core crystals⁵⁵. During data collection, the crystals were maintained at −10 °C. The crystals of the 440–525/GRE_{S4} complex belong to space group *P*2₁2₁2₁ and have cell dimensions *a* = 38.5 Å, *b* = 95.7 Å, *c* = 120.5 Å. Data were collected to 2.85 Å resolution with a Xuong–Hamlin area detector using the San Diego data collection software. A rotating Cu anode with a graphite monochromator was the X-ray source. Intensities were scaled with the program SCALEPACK (Z.O.). Heavy atom positions and disorder parameters were refined with the program PHARE-MAXLIKE (Z.O., Daresbury proceedings, in the press). Forty water molecules are included in the final model. The 16 residues of the carboxy termini are poorly ordered in both protein subunits and were not included in the model. Crystals of the 440–525/GRE_{S3} complex belong to space group *P*2₁2₁2₁ and have cell dimensions *a* = 38.5 Å, *b* = 99.7 Å, *c* = 121.9 Å. A molecular replacement solution was found using X-PLOR. The correlation function between *I*_{calc} and *I*_{obs} for the optimal fit was 61%.

* Figure of merit = $\int P(\theta) \exp(i\theta) d\theta / \int P(\theta) d\theta$, where *P* is the probability distribution of θ , the phase angle.
† $\sum |F_h^c| / \sum [|F_h^o| \exp(i\theta_c) + F_h^c] - |F_h^c|$, where $|F_h^c|$ is the calculated diffraction amplitude of the heavy atom, $|F_h^o|$ and $|F_h^c|$ are the observed amplitudes for the protein and heavy-atom derivative, respectively, θ_c is the calculated phase, and the sum is over all observations.

‡ Indicated by the position of substitution in the GRE_{S4} sequence. The bromine or iodine atom is at position 5 of uracil.

§ Includes 5934 Bijvoet differences.

(refs 25, 26); that is, binding of the first subunit favours the binding of the second. When an extra base pair is inserted between the hexameric half-sites as in GRE_{S4}, two monomers will still bind, but non-cooperatively (ref. 47, and L.P.F., unpublished results). The GRE_{S4} crystal structure suggests that the loss of cooperativity results from one subunit being forced to face a sequence to which it binds nonspecifically. That the protein remains the same symmetrical dimer for the GR/GRE_{S4} and GR/GRE_{S3} complexes indicates that the stability of the subunit interface must exceed the increment in stability provided by specific base interactions of a half-site. The subunits would otherwise be pulled apart and bind to the two half-sites of GRE_{S4} independently.

Discussion

Specificity of half-site sequence. Generally, the nuclear receptors have weak target affinity²⁶, high amino-acid sequence homology, and similar DNA response elements. How then are they targeted to their appropriate genes? The four classes of hormone response elements differ principally at positions 5 and 6 (Fig. 1b), and it is here where much of the discrimination must occur. For example, T(5)T(6) of the GRE becomes A(5)C(6) in the oestrogen response element (Fig. 1b). Mutagenesis experiments have shown that the glucocorticoid and oestrogen receptors are targeted to their respective response elements mostly by the amino-acid side chains at positions 458, 459 and 462 (glucocorticoid receptor numbering, for clarity; refs 9–12). These residues are Gly, Ser and Val, respectively, in the glucocorticoid receptor, and Glu, Gly, Ala in the oestrogen receptor. These observations may be rationalized by the specific side of the 440–525/GRE_{S4} crystal structure. First, the contact between the Val side chain and the 5-methyl group of T(5) stabilizes the 440–525/GRE_{S4} complex (Fig. 5b), so Val would be preferred over Ala and T over A in targeting the glucocorticoid receptor to a GRE. Second, Gly at 458 does not interact with any base, but Glu at this position could accept hydrogen bonds from the amino group of C(6) in the major groove, thus stabilizing an oestrogen receptor/ERE complex. The effect of substituting Gly for Ser 459 is less clear and may have indirect structural consequences. An

understanding of these changes will require structural analysis of the oestrogen receptor complex.

The role of the spacer in the GRE. Some hormone response elements have a three-base-pair spacer like the GRE and ERE, whereas others, such as the thyroid (Fig. 1b), retinoic acid, vitamin D3 and related receptors, do not⁴³. Presumably, either these receptors do not dimerize, or their DNA-binding domains interact in some way different from that described here for 440–525.

For the spacer in a hormone response element to act as a discriminating feature of the DNA target sequence, the protein subunits must dimerize in complementary register with the target half-sites. This requires that the stability of the dimer interface be comparable to, or greater than, the difference in affinity between a specific and nonspecific interaction at a receptor's half-site. The surprising feature of the glucocorticoid receptor/GRE interaction (and the analogous oestrogen receptor/ERE) is that the DNA-binding domain itself has little or no capacity to dimerize^{20,21,24}. But when the receptor binds to DNA, the target-DNA acts as an allosteric activator of recognition by providing a scaffold to bind the subunits in the correct juxtaposition for dimerization and by stabilizing substructures that support a strong dimer interface. Thus, the cooperative enhancement of DNA binding and dimerization favours binding to targets with appropriately spaced half-sites.

Recognition of diverse GREs. Among naturally occurring GREs, none adheres strictly to the idealized palindromic sequence⁴⁸.

TABLE 2 Refinement summary

Resolution (Å)	6.5–4.9	4.1	3.6	3.2	2.9	Total
Predicted	1,138	1,408	1,622	2,128	2,456	8,752
Refined	1,072	1,352	1,631	1,986	1,628	7,669
R-factor* (%)	20.5	19.6	19.1	18.9	19.6	19.6

* R-factor = $\sum |F_o - F_c| / \sum |F_o|$. The root-mean-square deviation of bond lengths is 0.017 Å; the r.m.s. deviation of interbond angles is 2.4°; and the r.m.s. shifts on the last refinement cycle for the positions and B values is 0.011 Å and 0.18 Å², respectively.

Indeed, the receptor binds with graded affinities to a broad range of sequences differing in their adherence to the consensus. Our results indicate that the receptor recognizes different sequences not by adjustments in its dimeric structure, but rather by making different types of DNA contacts. Thus, the protein binds in a half-specific/half-nonspecific mode with GRE_{S4} and, by analogy, with naturally occurring GREs that display only one clear half-site⁴. This implies that the GRE_{S4} and GRE_{S3} represent two extremes of a spectrum of selective interactions between the glucocorticoid receptor and naturally occurring GREs. These findings also provide a basis for considering how the receptor, perhaps in association with other factors, might interact with GREs bearing little resemblance to the consensus sequence^{49,50}.

The role of the α helix. The amino-terminal helix of 440–525 is another example of an α helix that provides most of the specifying base contacts of a transcriptional regulatory protein. As found for the helix–turn–helix motif of prokaryotic regulators,

the contact surface is stabilized and oriented by a well conserved, local constellation of side chains from other parts of the protein (the Zn centre and hydrophobic core here; the hydrophobic brace in the helix–turn–helix⁵¹). Another common feature is that a pair of helices, one from each subunit of a dimer, is symmetrically disposed in successive major grooves of the target DNA. This is in contrast to the helix–turn–helix of the *engrailed* homeodomain, where neither the protein nor the target are dyad symmetric, and the elongated second helix starts and ends outside the major groove⁵².

Zn coordination has more than one functional role in 440–525: it stabilizes not only the helices at the carboxy-terminal end of both fingers, but also the fold of the carboxy-terminal module that is essential for dimerization. The use of Zn²⁺ to nucleate substructures in the folding motifs of many transcription factors may be an evolutionary reflection of the structural economy and versatility provided by metal coordination. □

Received 9 April; accepted 26 June 1991.

- Chandler, V. L., Maier, B. A. & Yamamoto, K. R. *Cell* **33**, 489–499 (1983).
- Ponta, H., Kennedy, N., Skroch, P., Hynes, N. E. & Groner, B. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1020–1024 (1985).
- Issemann, I. & Green, S. *Nature* **347**, 645–649 (1990).
- Beato, M. *Cell* **56**, 335–344 (1988).
- Evans, R. M. *Science* **242**, 889–895 (1988).
- Nauber, U. *et al.* *Nature* **336**, 489–492 (1988).
- Segraves, W. A. & Hogness, D. S. *Genes Dev.* **4**, 204–219 (1990).
- Wang, L.-H. *et al.* *Nature* **340**, 163–166 (1989).
- Danielson, M., Hinck, L. & Ringold, G. M. *Cell* **57**, 1131–1138 (1989).
- Mader, S., Kumar, V., de Verneuil, H. & Chambon, P. *Nature* **338**, 271–274 (1989).
- Umesono, K. & Evans, R. M. *Cell* **57**, 1139–1146 (1989).
- Green, S., Kumar, V., Theulaz, I., Wahli, W. & Chambon, P. *EMBO J.* **7**, 3037–3044 (1988).
- Freedman, L. P. *et al.* *Nature* **334**, 543–546 (1988).
- Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **4**, 1609–1614 (1985).
- Brown, R. S., Sander, C. & Argos, P. *FEBS Lett.* **186**, 274–274 (1985).
- Berg, J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 99–102 (1988).
- Klug, A. & Rhodes, D. *Trends mol. Sci.* **12**, 464–469 (1987).
- Klevit, R. E., Herriott, J. R. & Horvath, S. J. *Proteins* **7**, 215–226 (1990).
- Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. *Science* **245**, 635–637 (1989).
- Hard, T. *et al.* *Science* **249**, 157–160 (1990).
- Schwabe, J. W. R., Neuhaus, D. & Rhodes, D. *Nature* **348**, 458–461 (1990).
- Pavletich, N. P. & Pabo, C. O. *Science* **252**, 809–817 (1991).
- Klock, G., Strahle, U. & Schutz, G. *Nature* **329**, 734–736 (1987).
- Freedman, L., Yamamoto, K., Luisi, B. F. & Sigler, P. *Cell* **54**, 444 (1988).
- Tsai, S. Y. *et al.* *Cell* **55**, 361–369 (1988).
- Hard, T., Dahlman, K., Carlstedt-Duke, J., Gustafsson, J.-A. & Rigler, R. *Biochemistry* **29**, 5358–5364 (1990).
- Schevitz, R. W., Podjarny, A. D., Zwick, M., Hughes, J. J. & Sigler, P. B. *Acta crystallogr.* **A37**, 669–677 (1981).
- Wang, B.-C., *Meth. Enzym.* **115**, 90–112 (1985).
- Finzel, B. C. *J. appl. Crystallogr.* **20**, 53–57 (1987).
- Westof, E., Dumas, P. & Moras, D. *J. molec. Biol.* **184**, 119–145 (1985).
- Hendrickson, W. A. & Konnert, J. H. in *Structure, Conformation, Function and Evolution* Vol. 1 (ed. Srinivasa, R.) 43–57 (1981).
- Otwinski, Z. *et al.* *Nature* **335**, 321–329 (1988).
- Bringer, A. *Crystallographic Computing 4: Techniques and New Technologies* (ed. Isaacs, N. W. & Taylor, M. R.) (Oxford University Press, 1988).
- Saenger, W. *Principles of Nucleic Acid Structure* 120 (Springer-Verlag, New York, 1984).
- Somers, W. A. thesis, Univ. Leeds, UK (1991).
- Schultz, S. & Steitz, T. A. *Science* (in the press).
- Schena, M., Freedman, L. P. & Yamamoto, K. R. *Genes Dev.* **3**, 1590–1601 (1989).
- Severne, Y., Wieland, S., Schaffner, W. & Rusconi, S. *EMBO J.* **7**, 2503–2508 (1988).
- Summers, M. F., South, T. L., Kim, B. & Hare, D. R. *Biochemistry* **29**, 329–340 (1990).
- Picard, D. & Yamamoto, K. R. *EMBO J.* **6**, 3333–3340 (1987).
- Scheidereit, C. & Beato, M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3029–3033 (1984).
- Scheidereit, C., Westphal, H. M., Carlson, C., Bosshard, H. & Beato, M. *DNA* **5**, 383–391 (1986).
- Forman, B. M. & Samuels, H. H. *Molec. Endocrin.* **4**, 1293–1301 (1990).
- Hughes, M. R. *et al.* *Science* **242**, 1702–1705 (1988).
- Rhodes, D. & Klug, A. *Nature* **286**, 573–578 (1980).
- Dahlman-Wright, K., Wright, A., Gustafsson, J.-A. & Carlstedt-Duke, J. *J. biol. Chem.* **266**, 3107–3112 (1991).
- Dahlman-Wright, K., Sitala-Roos, H., Carlstedt-Duke, J. & Gustafsson, J.-A. *J. biol. Chem.* **265**, 14030–14035 (1990).
- Jantzen, H.-M. *et al.* *Cell* **49**, 29–38 (1987).
- Sakai, D. D. *et al.* *Genes Dev.* **2**, 1144–1154 (1988).
- Diamond, M. I., Miner, J. N., Yoshinaga, S. K. & Yamamoto, K. R. *Science* **249**, 1266–1272 (1990).
- Zhang, R.-G. *et al.* *Nature* **327**, 591–597 (1987).
- Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. *Cell* **63**, 579–590 (1990).
- Read, R. J. *Acta crystallogr.* **A42**, 140–149 (1986).
- Presley, J. P. *J. appl. Crystallogr.* **21**, 572–576 (1988).
- Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. *Nature* **311**, 532–536 (1984).

ACKNOWLEDGEMENTS. We thank R. Basavappa, N. Duke, R. Hegde, P. Pepin, S. White, S. Luisi and R.-g. Zhang for help in collecting preliminary data at Cornell synchrotron and for encouragement and support, J. Flory (Howard Hughes Medical Institute, Yale) for synthesis of oligonucleotides. Work at Yale was supported by the USPHS, and work at UCSF by the USPHS and the NSF. B.L. was supported in part by the Jane Coffin Childs Memorial Fund for Cancer Research, and L.P.F. by the NIH and the Bank of America–Giannini Foundation. The coordinates of the 440–525/DNA complex have been submitted to the Brookhaven Protein Data Bank.

Gravitationally unbound comets move in predominantly retrograde orbits

John J. Matese, Patrick G. Whitman
& Daniel P. Whitmire

Department of Physics, The University of Southwestern Louisiana,
Lafayette, Louisiana 70504-4210, USA

COMETS are presumed to enter the inner Solar System from the Oort cloud, a repository of comets more than 10^4 AU from the Sun. Provided the perturbing effects of planetary encounters are taken into account, the original orbital energy of a comet can be calculated, and is negative or positive according to whether the comet's orbit is bound or unbound. The Oort effect¹ is the tendency for the original energies of long-period (>200 -yr) comets to fall within a narrow range: about 25% of such comets have original energies in the upper 0.2% of the total energy range. In addition, 10% of long-period comets are unbound, and it has been found² that positive original energy correlates with distance of closest approach to the Sun. We report here a further correlation: unbound comets are more likely to move in retrograde orbits. We suggest that this anomaly comes about because of the omission from the orbital energy determination of non-gravitational effects arising from enhanced volatility.

The orbit of a comet may be bound (elliptic) or unbound (hyperbolic). Its dimensions are characterized by the semimajor axis, a , and the orbital energy is proportional to $-1/a$. For the Earth, $a = 1$ AU $\approx 1.5 \times 10^8$ km. Oort concluded from the evidence of 19 well-determined comet orbits that the Sun was surrounded by a comet cloud having reciprocal semimajor axes in the range $0 < 1/a < 100$ units (1 unit = 10^{-6} AU⁻¹); in other words, a is at least 10^4 AU. Subsequent analyses^{2,3} have narrowed that range considerably.

One widely accepted model of comet-cloud evolution⁴ considers unaccreted comets in the region of Uranus and Neptune being pumped by these planets into more extended orbits. When semimajor axes grow to $\sim 5,000$ AU, galactic tidal torques change the angular momentum, thereby moving some of the points of closest approach (perihelia) beyond the planets. Eventually, passing molecular clouds and stars further diffuse and randomize the orbits to populate the inner cloud⁵ and the outer (Oort) cloud. The dominant mechanism that makes the Oort cloud observable, but the contiguous inner cloud largely unobserved, is the galactic tidal torque⁶⁻¹³ which causes changes in the perihelion distance proportional to the seventh power of semimajor axis. The distribution of comet flux also depends strongly on a , decreasing rapidly with increasing semimajor axis. Given that a convolution of these two functions, both steeply dependent on a , determines the probability of a comet entering the inner Solar System from the Oort cloud, one can explain the extraordinary narrow range of semimajor axis and energy for the observed Oort cloud⁹.

While transcribing data from Marsden's catalogue¹⁴ in an analysis of the effects of galactic tidal torques on cometary orbits, we became intrigued by the data shown in Table 1. It lists those hyperbolic comets whose original orbit determinations have been characterized² as class I (high quality) or II (intermediate quality). Also listed is the formal measured error δ , the orbit inclination angle, i , and the perihelion distance, q . One observes that 14 of the 15 most hyperbolic comets are retrograde. Even at a formal measured error of 5δ , seven of these comets (1895 IV, 1899 I, 1953 II, 1955 V, 1957 III, 1960 II and 1989 XIX) remain hyperbolic, and all are retrograde with $q < 0.9$ AU.

Here we aim to determine whether one can confidently correlate any deviation from the Oort tendency with orbital param-

eters. This in turn might suggest an explanation for the deviation. We conservatively restrict our consideration to the 75 new class-I comets having $1/a < 100$ units ($a > 10^4$ AU). The class I designation is biased against small perihelia distances² and thus considerably reduces the data base containing the effects that we wish to study. We comment on this later.

The first step in our investigation is to subdivide the class-I set into prograde (direct) and retrograde subsets and to list the paired values ($1/a$, δ^2) in order of ascending q . To emphasize the more accurately determined orbits, each value is given the relative weight δ^{-2} , an appropriate choice for minimizing chi-squared¹⁵. We create cumulative weighted mean arrays ($\langle 1/a \rangle$, $\langle \delta^2 \rangle$) where each member of the array is the mean of all previous values having $q < q_c$, a specified value of q . We then range q_c from 0.5 AU to 4.0 AU in steps of 0.1 AU. In addition to the results for $q < q_c$, we obtain the mean for the complement set $q > q_c$ and the cumulative weighted standard deviation (or scatter) about the mean, σ .

Figure 1a shows the mean for the original reciprocal semimajor axis, Fig. 1b the scatter and Fig. 1c the r.m.s. formal measured error. These presentations are variants of those given in Marsden *et al.*^{2,3}. Horizontal dashed bars in Fig. 1a delineate the range $1/a = 36 \pm 14$ (1 s.d.), which is the weighted fit to the data for $q > 2$ AU and is our adopted definition for the observed Oort population. The essential point to be extracted from Fig. 1a is that almost all of the deviation from the Oort tendency is contained in the small- q retrograde population. In contrast, the small- q prograde population is completely normal. Figure 1b further emphasizes this observation, as it shows the observed scatter about the mean for the four subgroups. The galactic tide is capable of making Oort comets with $5 < 1/a < 50$ observable in a single orbit⁶⁻¹¹. Individual comets whose semimajor axes are outside this range require an alternative explanation.

Sharp changes in the retrograde $q < q_c$ curves at $q_c = 1.2$ AU are due to the heavily weighted comet 1987 VII ($1/a = 67$, $\delta = 2$)

TABLE 1 Hyperbolic comets

Name	$1/a$	Class	δ	$\cos(i)$	q
1976 I	-734	II	167	-0.070	0.864
1955 V	-727	II	121	-0.301	0.885
1959 III	-446	II	524	0.975	1.251
1989 XIX	-218	II	34	-0.003	0.642
1895 IV	-172	I	8	-0.784	0.192
1971 V	-142	II	40	-0.337	1.233
1960 II	-135	II	23	-0.937	0.504
1953 II	-125	I	9	-0.125	0.778
1940 III	-124	II	316	-0.683	1.062
1986 XVII	-121	II	76	-0.115	0.921
1899 I	-109	I	9	-0.832	0.327
1957 III	-98	I	6	-0.498	0.316
1968 VI	-82	II	116	-0.801	1.160
1904 II	-75	II	98	-0.167	1.882
1911 IV	-74	II	47	-0.113	0.303
1898 VIII	-71	I	36	0.924	2.285
1932 VII	-56	I	15	0.201	1.647
1975 XI	-56	II	34	0.332	0.219
1942 VIII	-34	I	13	-0.991	4.113
1892 VI	-27	I	11	0.908	0.976
1981 V	-26	II	137	-0.625	2.112
1849 II	-25	II	38	0.388	1.159
1886 I	-18	I	8	0.129	0.642
1983 XVI	-18	I	11	-0.511	2.255
1946 I	-13	I	3	0.296	1.724
1988 II	-4	I	16	0.230	3.333
1947 I	-1	I	5	-0.312	2.408

Orbit data for the set of hyperbolic comets deemed class I or class II. The original value of reciprocal semimajor axis, $1/a$, and the formal measured error in the osculating value of reciprocal semimajor axis, δ , are given in units of 10^{-6} AU⁻¹. Also shown are the orbit inclination to the ecliptic, i , and the perihelion distance, q (AU).

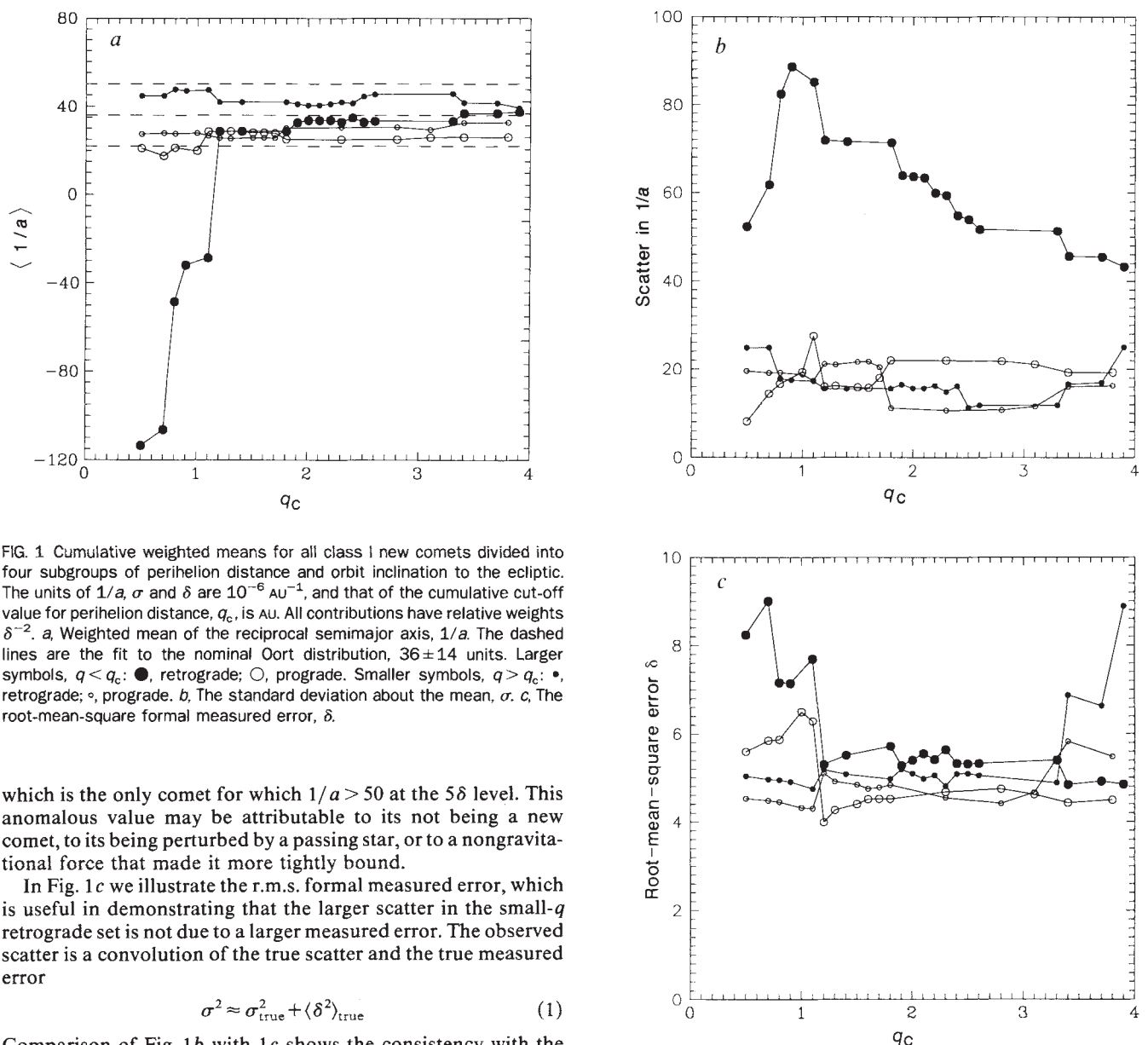


FIG. 1 Cumulative weighted means for all class I new comets divided into four subgroups of perihelion distance and orbit inclination to the ecliptic. The units of $1/a$, σ and δ are 10^{-6} AU $^{-1}$, and that of the cumulative cut-off value for perihelion distance, q_c , is AU. All contributions have relative weights δ^{-2} . *a*, Weighted mean of the reciprocal semimajor axis, $1/a$. The dashed lines are the fit to the nominal Oort distribution, 36 ± 14 units. Larger symbols, $q < q_c$: $\bullet$, retrograde; $\circ$, prograde. Smaller symbols, $q > q_c$: $\bullet$, retrograde; $\circ$, prograde. *b*, The standard deviation about the mean, σ . *c*, The root-mean-square formal measured error, δ .

which is the only comet for which $1/a > 50$ at the 5δ level. This anomalous value may be attributable to its not being a new comet, to its being perturbed by a passing star, or to a nongravitational force that made it more tightly bound.

In Fig. 1c we illustrate the r.m.s. formal measured error, which is useful in demonstrating that the larger scatter in the small- q retrograde set is not due to a larger measured error. The observed scatter is a convolution of the true scatter and the true measured error

$$\sigma^2 \approx \sigma_{\text{true}}^2 + \langle \delta^2 \rangle_{\text{true}} \quad (1)$$

Comparison of Fig. 1b with 1c shows the consistency with the suggestion² that the true measured error may be up to 3 times larger than the formal value. We suggest adopting as the true scatter in the Oort cloud the value $\sigma_{\text{true}} \approx 10 \pm 4$ units.

We comment now on the confidence level of the assertion that the small- q retrograde population has a different distribution from the remaining comets used to define the observed Oort population $1/a = 36 \pm 14$ (formal r.m.s. $\delta = 5$). An appropriate statistic for comparing a sample mean with a population mean is the t statistic^{15,16}

$$t = (\langle 1/a \rangle - 36) / (\sigma / N^{1/2}) \quad (2)$$

where N is the number of comets in the sample. An illustrative calculation is for the data at $q_c = 0.7$ AU, where $\langle 1/a \rangle = -107$, $\sigma = 62$ and $N = 5$, yielding $t = -5.2$. For a one-sided t -test with four degrees of freedom, the confidence level is 99.7% that the difference between this sample mean and the Oort value is not due to random chance.

Another statistic of interest is the F statistic, which compares a sample variance, σ^2 , with a population variance¹⁵

$$F = (\sigma^2 / \langle \delta^2 \rangle) / (14^2 / 5^2). \quad (3)$$

For $q_c = 0.7$ AU, we find $\langle \delta^2 \rangle = 81$ so that $F = 6.0$. Again the confidence level exceeds 99% that the sample variance does not differ from the Oort value because of random chance. The results

do not imply that all small- q retrograde comets should be non-Oort-like, nor do they imply that all non-Oort-like comets should be small- q retrograde. We realise that one could object that the above effects are dominated by a relatively small number of comets. The correlation described is also seen, however, when the data base of class II comets alone is analysed. It is also comparably significant if uniform weighting is adopted.

We now consider possible explanations for the observed correlation. A selection effect that correlates retrograde orbits with small q can be rejected by analysis. A selection effect that correlates unusually large measurement errors with the small- q retrograde population can be rejected by inspection of Fig. 1c. Any injection mechanism for this population that is different from the galactic tide would need to explain the small- q preponderance. Neither stellar impulses⁵ nor an interstellar origin can provide the explanation. The hyperbolic character of the orbits shows that they have not been previously perturbed by the planets.

We are then left with the possibility that deviations from the Oort tendency are caused by enhanced nongravitational forces which correlate with orbits that are preferentially small- q (as suggested in ref. 2) and retrograde. A dynamical parameter

suggested is the energy of relative motion between a comet and a meteoroid in a prograde ecliptic circular orbit of radius r ,

$$U^2/2 = (3GM_\odot/2r)[1 - (8q/9r)^{1/2} \cos(i)] \quad (4)$$

If $r = q$, the energy ratio for comets of $i = 180^\circ$ to those of $i = 0^\circ$ is 34 (a collision between a retrograde comet and a prograde meteoroid will have more effect than a similar collision between prograde comet and prograde meteoroid). We suggest that processing of the comet's mantle by an ecliptic population of meteoroids is the cause of enhanced volatility with an accompanying deviation from the Oort clustering. The known

meteoroid population¹⁷⁻¹⁹ would directly expose only a small fraction of the volatiles underlying a mantle ($\sim 10^{-5}$ – 10^{-4} for the anomalous comets if they had 1-cm mantles), and crater growth needs to be demonstrated. Alternatively, impacts of an as-yet-undiscovered population of larger objects (~ 10 m), previously invoked as a possible explanation for erratic behaviour²⁰, may be the cause. Six of the seven anomalous comets with trajectories that are hyperbolic at the 5 σ level were observed to split physically, or to be made more Oort-like in their energies when the effects of nongravitational forces were modelled (B. G. Marsden, personal communication). The exception is comet 1953 II, for which the modelling is inconclusive². $\square$

Received 22 January; accepted 21 June 1991.

1. Oort, J. H. *Bull. astr. Inst. Neth.* **11**, 91–110 (1950).
2. Marsden, B. G., Sekanina, Z. & Everhart, E. *Astr. J.* **83**, 64–71 (1978).
3. Marsden, B. G. & Sekanina, Z. *Astr. J.* **78**, 1118–1124 (1973).
4. Duncan, M., Quinn, T. & Tremaine, S. *Astr. J.* **94**, 1330–1338 (1987).
5. Hills, J. G. *Astr. J.* **86**, 1730–1740 (1981).
6. Byl, J. *Earth Moon Planets* **36**, 263–273 (1986).
7. Heisler, J. & Tremaine, S. *Icarus* **65**, 13–26 (1986).
8. Heisler, J., Tremaine, S. & Alcock, C. *Icarus* **70**, 269–288 (1987).
9. Matese, J. J. & Whitman, P. G. *Icarus* **82**, 389–401 (1989).
10. Morris, D. E. & Muller, R. A. *Icarus* **65**, 1–12 (1986).
11. Torbett, M. V. *Mon. Not. R. astr. Soc.* **223**, 885–895 (1986).
12. Yabushita, S. *Mon. Not. R. astr. Soc.* **231**, 723–733 (1988).

13. Yabushita, S. *Astr. J.* **97**, 262–264 (1989).
14. Marsden, B. G. *Catalogue of Cometary Orbits* 6th edn (Smithsonian Astrophysical Observatory, Cambridge, Massachusetts, 1989).
15. Meyer, S. L. *Data Analysis*, 254–290 (Wiley, New York, 1975).
16. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes*, 464–497 (Cambridge University Press, 1986).
17. Leinert, C. I., Richter, I., Pitz, E. & Planck, B. *Astr. Astrophys.* **103**, 177–188 (1981).
18. Leinert, C. I., Roser, S. & Buitrago, J. *Astr. Astrophys.* **118**, 345–357 (1983).
19. Grün, E., Zook, H. A., Fechtig, H. & Giese, R. H. *Icarus* **62**, 244–272 (1985).
20. Marsden, B. G. & Sekanina, Z. *Astr. J.* **76**, 1135–1151 (1971).

ACKNOWLEDGEMENTS. We thank M. Bailey, A. Jackson and B. G. Marsden for suggestions. This work was partially supported by a NASA-Ames Research Center University Consortium grant.

Effect of cyanide on the photoelectrochemical response of n-CdSe/Fe(CN)₆^{4-/3-} electrolytic cells

Gayatri Seshadri, Jonathan K. M. Chun & Andrew B. Bocarsly*

Department of Chemistry, Frick Laboratory, Princeton University, Princeton, New Jersey 08544-1009, USA

PHOTOELECTROCHEMICAL cells based on cadmium chalcogenide electrodes in an aqueous ferrocyanide electrolyte¹⁻⁹ provide a model system for studying the fundamental processes associated with photo-induced interfacial charge transfer, and also hold the promise of converting solar energy to electricity and useful chemical products with high efficiency. Licht⁹ has suggested that it is possible to devise n-CdSe/[Fe(CN)₆^{3-/4-}]_{aq} cells with a maximum solar-energy conversion efficiency of 16.4%, which derives from an open-circuit photovoltage of ~ 1.2 V, a value that approaches the thermodynamic limiting value for a semiconductor with a bandgap of 1.7 eV. Licht suggested that improvements in energy conversion efficiency are possible by modifying the electrolyte so as to prevent chemical modification of the semiconductor surface while optimizing the equilibrium concentrations of critical solution species. Here we report that addition of KCN to a ferro/ferricyanide electrolyte does not prevent surface modification on either the (11 $\bar{2}$ 0) face of n-CdX (where X is S or Se) or the (0001) face of n-CdS. We show instead that the current-voltage response observed by Licht is due to a hidden negative ΔG in the system associated with the thermodynamically 'downhill' oxidation of CN⁻, which accounts for as much as 700 mV of the reported 'open circuit' photovoltage.

Previous studies⁴⁻⁸ have demonstrated the formation of a layer of [CdFe(CN)₆]^{2-/1-} on the cadmium-rich (0001) surface of n-CdX-based photoelectrochemical cells and the importance of this surface modification in stabilizing such cells against photo-

induced anodic dissolution of the semiconducting electrode. For n-CdS-based cells, this surface layer considerably increases the optical energy conversion capability. Diffuse-reflectance Fourier transform infrared (FTIR) spectroscopy of electrode surfaces that have been used in illuminated photoelectrochemical cells provides direct evidence that such an overlayer is obtained on n-CdSe on photolysis (in an electrolyte composed of a mixture of K₄[Fe(CN)₆] and K₃[Fe(CN)₆], (20:1) with 0.1 M KCN adjusted to pH 13.7 with KOH⁹). Typical spectroscopic results are shown in Fig. 1 for an electrode irradiated at 488 nm (20 mW cm⁻²) for 1 h at 0.0 V relative to the standard

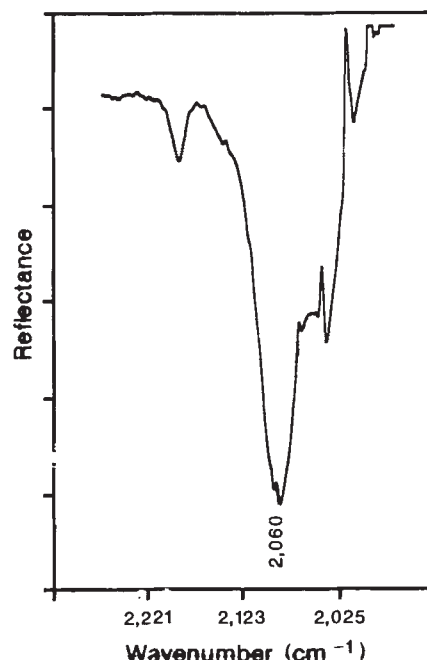
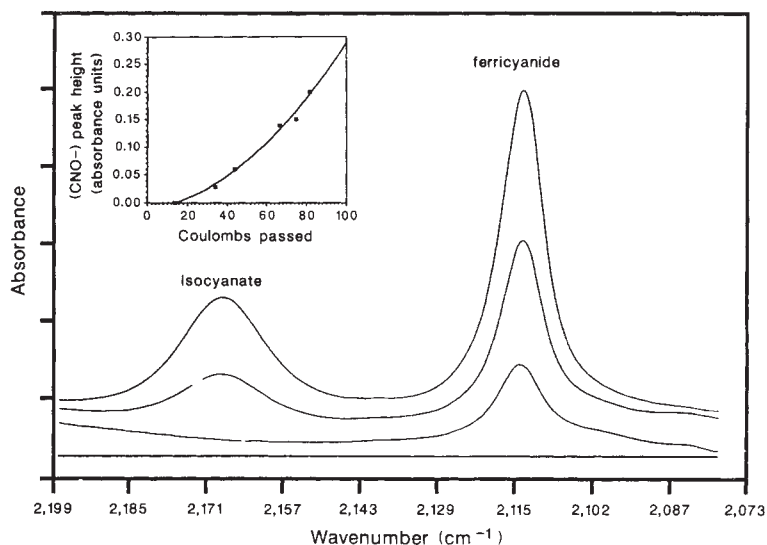


FIG. 1 Diffuse-reflectance FTIR spectrum of the (11 $\bar{2}$ 0) face of single-crystal n-CdSe at 0 V relative to SCE, after 20 mW cm⁻² irradiation from the 488 nm line of an argon ion laser for one hour. Before this, the electrode was polished with an alumina slurry and chemically etched with a 5% solution of bromine in methanol. All FTIR spectra were obtained using a Nicolet 730 or 800 FTIR spectrometer.

* To whom correspondence should be addressed.

FIG. 2 FTIR spectra of the electrolyte in the n-CdSe/0.25 M $K_4[Fe(CN)_6]$, 0.0125 M $K_3[Fe(CN)_6]$, 0.1 M KCN, 0.5 M KOH system at various times during a typical experimental run (bottom to top, 0, 1.2, 3.4 and 7.0 hours). Electrode pretreatment and irradiation are described in Fig. 1. The electrode, of area 0.25 cm², was held at 0 V relative to SCE in the above electrolyte for ~8.4 hours in a standard two-compartment cell with argon bubbling. Small aliquots were removed from the cell at regular intervals and pressed between CaF₂ plates separated by a 25×12×0.015 mm Teflon spacer. Each spectrum was measured relative to the spectrum recorded at time $t=0$. The split laser beam irradiated a control solution at the same intensity.

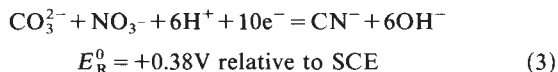
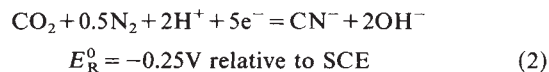


calomel electrode (SCE). The key feature in this spectrum is a transition centred at 2,060 cm⁻¹, is associated with a bridging cyanide, and indicates the presence of a cadmium ferricyanide species on the electrode surface⁵⁻⁷. Electrodes irradiated in an electrolyte containing only KCN did not yield this spectroscopic feature. As the spectrum is obtained in a reflectance measurement, it is difficult to quantify the amount of $[CdFe(CN)_6]^{2-/1-}$ present on the electrode surface. Qualitatively, it seems that addition of cyanide to the $[Fe(CN)_6]^{4-/3-}$ electrolyte decreases the amount of surface material. Nonetheless, a substantial cadmium cyanometallate layer is present when solutions containing 0.1 M KCN are used. Although control FTIR experiments in which an electrode with such a surface layer was washed with a concentrated KCN solution (~1 M) not containing ferro or ferricyanide indicate that such solutions do effectively remove the cyanometallate overlayer, this is not the experimental condition described by Licht⁹.

There are numerous reports¹⁰ on the oxidation of potassium cyanide at metal electrodes. The standard redox potential E_R^0 for the process



is reported to be -0.73V relative to SCE¹¹. Additionally, it has been reported that CN⁻ can be completely oxidized to CO₂ and NO₃⁻ or N₂ electrochemically as indicated below.



As the valence band edge of CdSe in $[Fe(CN)_6]^{4-/3-}$ (aq) electrolyte has been placed at ~+0.5V relative to SCE, we have looked into the possibility of oxidation of cyanide in the n-CdSe/ $K_4[Fe(CN)_6]$, $K_3[Fe(CN)_6]$, KCN, KOH system. Hinman *et al.*¹² reported that electrogenerated isocyanate formation from cyanide in aqueous solution (pH 12) could be monitored by observing an infrared absorbance centred at 2,169 cm⁻¹. We have reproduced this observation by adding analytical quantities of KCNO to the electrolyte of present interest. This addition produces a transition centred at 2,169 cm⁻¹ which obeys Beer's law (absorbance per mol per cm, $\epsilon = 7.3 \text{ M}^{-1} \text{ cm}^{-1}$). The FTIR spectrum of an aqueous electrolyte composed of 0.25 M $K_4[Fe(CN)_6]$, 0.0125 M $K_3[Fe(CN)_6]$ and 0.1 M KCN, pH adjusted with KOH, is shown in Fig. 2. Initially, no absorbance corresponding to CNO⁻ was observed. An

intense absorbance associated with the vibration frequency of CN, ν_{CN} , in $[Fe(CN)_6]^{4-}$ (omitted from Fig. 2 for clarity) is observed at 2,040 cm⁻¹. On photoelectrolysis at a n-CdSe electrode (a two-compartment cell was used to distinguish the products generated at the anode from the those at the cathode), an absorption band at 2,169 cm⁻¹ unambiguously indicates the formation of isocyanate (Fig. 2). Simultaneously, an absorption centred at 2,115 cm⁻¹ is observed to grow in. This transition is assigned to ν_{CN} in $[Fe(CN)_6]^{3-}$ indicating the concomitant oxidation of ferrocyanide. In an electrolysis time of 8.4 h, 85.5 coulombs of charge were passed, producing 16.2 mmol of CNO⁻ according to FTIR analysis. The inset to Fig. 2 shows the increase in isocyanate as a function of coulombs passed during the experiment. Control experiments in which the electrolyte is irradiated at the same intensity and for the same period of time in the absence of the photoanode demonstrate that no solution photochemistry occurs that produces CNO⁻.

From the data reported above, 26% of the total current or 15% of the photogenerated holes (based on the current doubling mechanism discussed below) are associated with cyanide oxidation. Cyanide oxidation therefore competes with ferrocyanide oxidation at the n-CdSe interface. This finding is rather

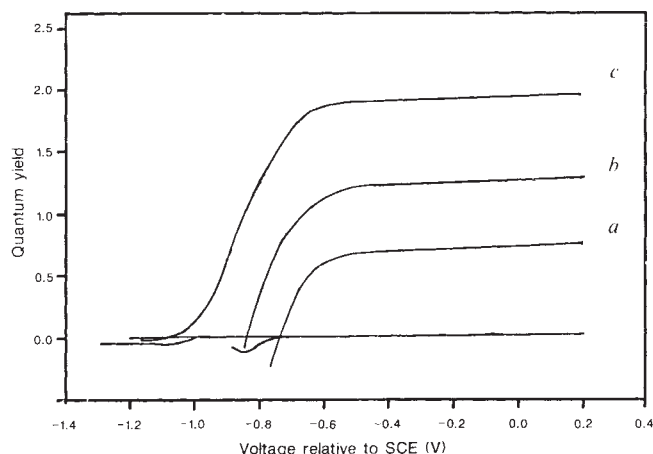


FIG. 3 Current-voltage curves (scanned at 10 mV s⁻¹) for n-CdSe (1120) in different electrolytes. a, 0.25 M $K_4[Fe(CN)_6]$, 0.0125 M $K_3[Fe(CN)_6]$, 0.5 M KOH. b, 0.25 M $K_4[Fe(CN)_6]$, 0.0125 M $K_3[Fe(CN)_6]$, 0.5 M KOH and 0.1 M KCN. c, 1 M KCN, 0.5 M KOH. Electrode pretreatment and irradiation are described in Fig. 1. A Princeton Applied Research 174A model potentiostat was used, and current-voltage curves were recorded using a Honeywell 320 XY recorder.

surprising when it is considered that cyanide oxidation is reportedly irreversible and that ferrocyanide oxidizes easily at many metal electrodes.

Typical steady-state current-voltage curves are shown in Fig. 3 for the n-CdSe/[Fe(CN)₆]^{4-/3-} cell with and without added KCN. The principal reported improvement in conversion efficiency on addition of CN⁻ to the electrolyte derives from the observed shift in the photocurrent onset potential to more negative values. But this improvement is offset by the very negative redox potential of the CN⁻ couple. If the calculated redox potential for cyanide oxidation (-0.73V relative to SCE) is used as the short-circuit potential, then an open-circuit photopotential of only ~500 mV is calculated for the cyanide-containing cell, considerably less than that reported for the n-CdSe cell based on pure [Fe(CN)₆]^{3-/4-} (ref. 7). This calculation overestimates the extent to which Licht's results exceed the true open-circuit potential, but as a mixed potential condition (concomitant ferrocyanide and cyanide oxidation) exists, it serves to point out that the actual open-circuit photopotential is considerably less than the reported 1.2 V. An alternative way of looking at this cell would be to describe it as having a 'virtual' battery (the cyanide/isocyanate couple) in series with the [Fe(CN)₆]^{3-/4-} photoelectrochemical couple. This battery, the thermodynamically downhill oxidation of CN⁻, offsets the observed current-voltage curves by its output potential (~200–250 mV). The addition of cyanide not only causes a negative shift in the onset potential, but also results in a substantial increase in the plateau current. This variation leads to an apparent enhancement in the cell energy conversion efficiency. The 'enhancement' cannot be due to an improvement in the interfacial photoelectrochemical oxidation of ferrocyanide, as this process occurs with a quantum yield approaching unity at the n-CdSe interface^{6,7}. The limiting quantum yield for electron flow measured at the plateau region of the steady-state current-voltage profile for the alkaline n-CdSe-ferro/ferricyanide-cyanide system is 1.27–1.34 when 488-nm illumination is used (20 mW cm⁻²); see Fig. 3 curve (b). These values indicate that a current-doubling mechanism exists. As ferro/ferricyanide is a well established one-electron redox couple⁷, the current-doubling process must be associated with the oxidation of cyanide to isocyanate. The data shown in Fig. 3 confirm this conclusion. A limiting quantum yield of 2 is obtained for curve (c) in an electrolyte containing only 1 M KCN, verifying that CN⁻ is oxidized at the n-CdSe surface by means of a current-doubling mechanism.

Under alkaline conditions, the oxidation of cyanide at a metal electrode has been described as a simultaneous two-electron process^{10,11}. Typically, current doubling at a semiconductor-electrolyte interface is understood in terms of a mechanism involving an initial photoinduced one-electron transfer of a hole from the semiconductor valence band to the electroactive species in solution to generate an electroactive species having a very negative redox potential. This material can then inject an electron into the semiconductor conduction band. Consistent with this mechanism, we find that the quantum yield for electron flow decreases as the light intensity is increased for a photoelectrochemical cell with pure cyanide electrolyte. The current-doubling data presented above strongly suggest that CN⁻ is oxidized in two one-electron processes, not in a single two-electron event. A similar result has been reported for the oxidation of cyanide at a n-TiO₂ photoelectrode¹³. This is best explained in terms of the formation of a cyanogen free radical following the primary photoinduced transfer of a hole from the valence band to the cyanide. This high-energy species could then react with water, followed by injection of a second electron into the conduction band to generate isocyanate. We are still studying the mechanistic details of the reaction. It should be noted that the presence of a [CdFe(CN)₆]^{2-/1-} overlayer does not preclude the direct reaction of cyanide with the CdSe surface, as the overlayer is known to be porous¹⁴. The observation of

current doubling provides evidence that CN⁻ is oxidized directly at the n-CdSe interface. A more realistic value for the maximum optical energy conversion efficiency, therefore, would be that obtained in the cyanide-free electrolyte (curve a of Fig. 3), namely 16.1% (at 488 nm).

In the light of these findings, one cannot say that the addition of KCN to the alkaline [Fe(CN)₆]^{3-/4-} system enhances the conversion efficiency. Although there is some enhancement in quantum yield due to current doubling, the chemical product of this reaction (CNO⁻) is of little practical use (it is not a possible chemical fuel) and does not allow for the formation of a regenerative cell. □

Received 19 October; accepted 28 May 1991.

1. Noufi, R., Tench, D. & Warren, L. F. *J. electrochem. Soc.* **128**, 2363 (1981).
2. Tenne, R. *J. electrochem. Soc.* **130**, 2163–2169 (1983).
3. Reichman, J. & Russack, M. A. *J. electrochem. Soc.* **131**, 796–798 (1984).
4. Rubin, H.-D., Humphrey, B. D. & Bocarsly, A. B. *Nature* **308**, 339–341 (1984).
5. Rubin, H.-D., Arent, D. J. & Bocarsly, A. B. *J. electrochem. Soc.* **132**, 523–524 (1985).
6. Rubin, H.-D. thesis, Univ. of Princeton (1986).
7. Arent, D. J. thesis, Univ. of Princeton (1987).
8. Chen, Y. thesis, Univ. of Princeton (1990).
9. Licht, S. & Peramunage, D. *Nature* **345**, 330–333 (1990).
10. Kuhn, A. T. in *Modern Aspects of Electrochemistry Vol. 8* (eds Bockris, J. O'M. & Conway, B. E.) (1972).
11. *Encyclopedia of Electrochemistry of the Elements Vol. 7* (ed. Bard, A. J.) 206–215 (1976).
12. Hinman, A. S., Kydd, R. A. & Cooney, R. P. *JCS Faraday Trans. 1* **82**, 3525–3534 (1986).
13. Frank, S. N. & Bard, A. J. *J. Am. chem. Soc.* **99**, 4667–4667 (1977).
14. Sinha, S., Amos, L. J., Schmidt, M. H. & Bocarsly, A. B. *J. electroanal. Chem.* **210**, 323 (1986).

ACKNOWLEDGEMENTS. This work was funded by the US Department of Energy, Division of Basic Energy Sciences.

Optical enrichment of a racemic chiral crystal by X-ray irradiation

Yasuko T. Osano, Akira Uchida & Yuji Ohashi

Department of Chemistry, Tokyo Institute of Technology,
2-12-1, Ookayama, Meguro-ku, Tokyo 152, Japan

GENERATION of optical activity in a racemic mixture of enantiomers, a process relevant to the origin of life and of naturally chiral compounds¹, requires chirality in the interaction responsible for the transformation or the environment in which it occurs. Without violating this strict requirement, we have found a way in which optical enrichment of a racemic mixture can occur without any chiral specificity in the mechanism. A racemic mixture of (1-cyanoethyl)piperidine)cobaloxime, which contains the chiral 1-cyanoethyl group, crystallizes in a chiral space group (*P*₂₁₂₁₂) in which *R* and *S* enantiomers occupy crystallographically distinct (diastereoisomeric) positions in the asymmetric unit cell. We have prepared single crystals of the *D* configuration. Irradiating the crystal with X-rays at 343 K causes racemization of the *S* enantiomer while the *R* enantiomer remains unaltered. This racemization is thought to be due solely to the different volume constraints on the two enantiomers in their different crystal environments. The result is that the number of *R* molecules is increased at the expense of *S*, and so the overall composition of the crystal changes from racemic to enriched in the *R* enantiomer.

Previously, chiral β-lactam has been formed from an achiral oxo amide on exposure to visible light using the chiral crystal environment². The chiral alkyl groups bonded to the cobalt atom in some cobaloxime (another name for 2,3-butanedione-dioximatocobalt(III)) complex crystals have been found to be racemized by X-ray exposure without degradation of crystallinity^{3–5}. For crystals with the chiral 1-cyanoethyl group, the racemization process of various crystals was divided into three modes and the reaction rate was found to be closely related to the free space around the reactive group (the reaction cavity).

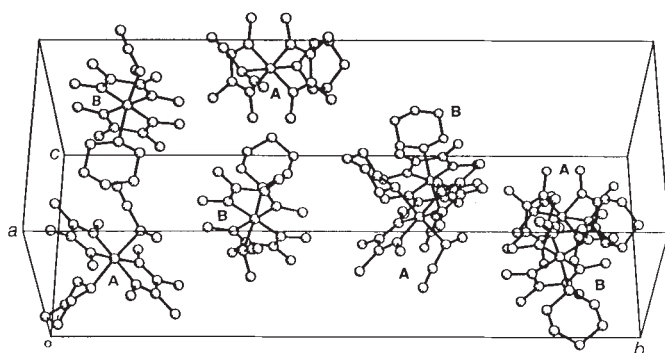


FIG. 1 Crystal structure of racemic (1-cyanoethyl)(piperidine) cobaloxime, (III). The absolute configurations of the cyanoethyl groups of molecules A and B are *R* and *S*, respectively.

We have recently⁶ prepared chiral crystals of (*R*-1-cyanoethyl)(piperidine)cobaloxime (structure I). X-ray crystal structure analysis showed that the crystal has two molecules, A and B, in its asymmetric unit. The chiral cyanoethyl group of molecule B was gradually changed to the disordered racemate by X-ray exposure at 333 K, although it remained unaltered at 293 K. The cyanoethyl group of molecule A was unchanged at either temperature. In the new crystal structure (II) produced by X-ray exposure, therefore, the ratio of *R* and *S* configurations of the cyanoethyl groups was 3:1. The difference in reactivity could be well explained by the different size of the reaction cavity around the cyanoethyl groups.

For comparison with (II), we have now obtained racemic crystals of (1-cyanoethyl)(piperidine)cobaloxime (structure III) from an aqueous methanol solution. The dark red crystals are orthorhombic with space group $P2_12_12_1$, and have cell dimensions $a = 11.699(3)$, $b = 30.70(1)$ and $c = 11.421(3)$ Å, and cell volume $V = 4102(1)$ Å³. These values are roughly the same as those of the crystals (I) and (II). We analysed the structure with the programs MULTAN78⁷ and SHELX76⁸ to give a final residual $R = 0.074$ for 2,759 reflections with $|F_0| > 3\sigma(|F_0|)$. The analysed structure (Fig. 1), to our surprise, is isomorphous to those of (I) and (II) except that molecule B has a *S*-1-cyanoethyl group instead of the *R*-1-cyanoethyl or disordered (*R,S*)-1-cyanoethyl group in (I) or (II). The molecular structures of A and B, crystallographically independent molecules, are almost the same, except for the configuration of the cyanoethyl group. Figure 2a shows the structure of B in the *S* configuration. As the crystals have two enantiomers, the above crystal was selected so that molecule A would have the same absolute configuration as that of (I). By seeding with finely powdered crystals of one or other enantiomer of (1-cyanoethyl)(piperidine)cobaloxime, crystals of the same absolute configuration are nucleated. This absolute configuration was confirmed by Bijvoet analysis using anomalous dispersion.

The crystal was not changed by X-ray irradiation at 293 K. When it was warmed to 343 K, however, the cell dimensions were gradually changed on exposure to X-rays. After a week, the crystal was cooled to 293 K. The final cell dimensions were as follows: $a = 11.717(3)$, $b = 30.771(5)$ and $c = 11.400(2)$ Å, with a final cell volume of $V = 4110(2)$ Å³. We then solved the crystal structure, collecting the intensity data and determining the structure as described for crystal (III) at room temperature. The final value of the residual R was 0.068 for 4,227 reflections. The structure analysis showed no change in molecule A, whereas significant peaks appeared around the cyanoethyl group of molecule B. These peaks were assigned to the cyanoethyl group with *R* configuration, and the occupancy factors of the newly appeared *R* and the original *S* cyanoethyl groups converged to the same value, 0.5. Figure 2b shows the disordered molecular structure of B with *S* and *R* configurations, which is very similar to that of molecule B in crystal (II).

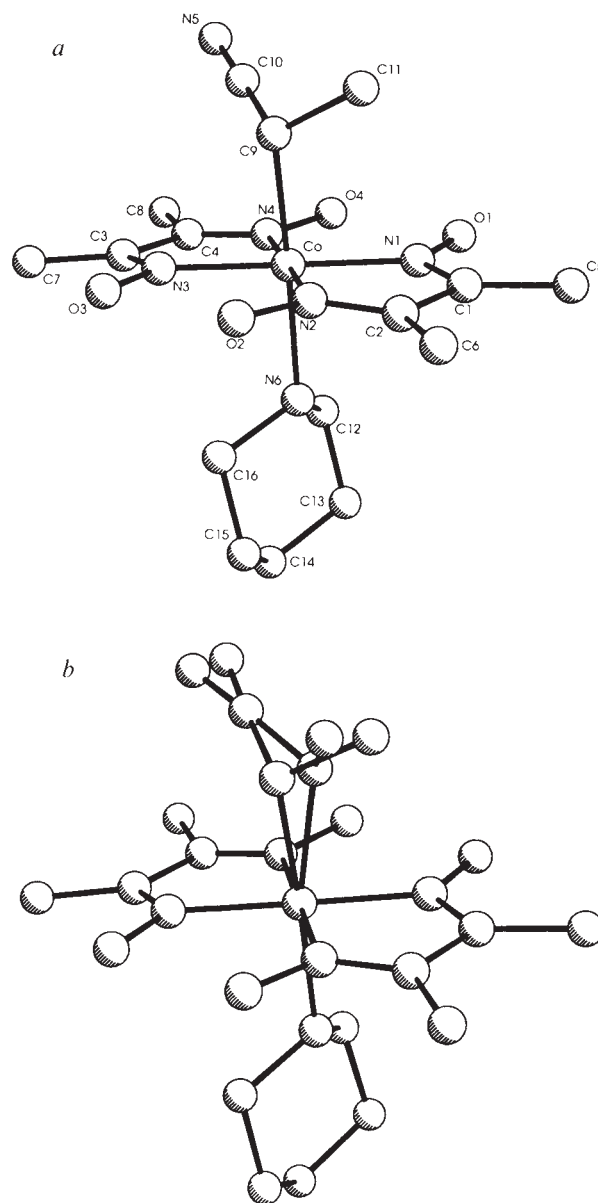


FIG. 2 a, Molecular structure of molecule B before the racemization. b, Molecular structure after one week's exposure to X-rays at 343 K. A small portion of the newly appeared *R*-cyanoethyl has been omitted for clarity.

For crystal (I), the volumes of the reaction cavities around the cyanoethyl groups of A and B were calculated to be 7.49 and 11.57 Å³ at 293 K, and 7.88 and 14.04 Å³ at 333 K, respectively⁶. For a cyanoethyl group isolated from the other groups in the crystal, the volume necessary for the racemization was estimated⁵ to be 11.5 Å³. These facts explain why only the B cyanoethyl group is converted to the disordered racemate at high temperatures. Although the volume of the reaction cavity for the present racemic crystals has not yet been calculated, the A and B cyanoethyl groups would occupy roughly the same volume as those in crystal (I). Only the B cyanoethyl group, therefore, would have enough space to take the disordered structure at 343 K. The entropy term should be a driving force for the racemization of the B cyanoethyl group on X-ray exposure in a unit cell nearly the same size as that of crystal (I). The A cyanoethyl group retains its configuration because it has insufficient free space to be able to racemize. Although the racemic-to-chiral conversion by X-ray exposure alone seems curious at first sight, it is a natural reaction from a thermodynamical point of view. □

Received 2 May; accepted 28 June 1991.

1. Addadi, L. & Lahav, M. *Origin of Optical Activity in Nature* (ed. Walker, D. C.) (Elsevier, New York, 1979).
2. Sekine, A., Hori, K., Ohashi, Y., Yagi, M. & Toda, F. *J. Am. chem. Soc.* **111**, 697–699 (1989).
3. Ohashi, Y. & Sasada, Y. *Nature* **267**, 142–144 (1977).
4. Ohashi, Y., Yanagi, K., Kurihara, T., Sasada, Y. & Ohgo, Y. *J. Am. chem. Soc.* **103**, 5805–5812 (1981).
5. Ohashi, Y. *Acc. Chem. Res.* **21**, 268–274 (1988).
6. Osano, Y. T. et al. *Acta crystallogr.* B47 (in the press).
7. Main, P. et al. *MULTAN78. A System of Computer Programs for the Automatic Solution of Crystal Structure from X-Ray Diffraction Data* (Universities of York and Louvain, 1978).
8. Scheelrick, G. M. *SHELX76. A Program for Crystal Structure Determination* (University of Cambridge, 1976).

Autorotation of many-sided bodies in an airstream

B. W. Skews

School of Mechanical Engineering, University of the Witwatersrand, Johannesburg, South Africa

AUTOROTATION is the rotation of an object in an airstream in the absence of any other driving force¹. It is important in the fall characteristics of some tree fruits, the growth of hailstones, and the trajectories of objects separating from aircraft and spacecraft. Previous studies have concentrated on thin, flat plates rotating about an axis normal to the airstream, but in many applications—those relating to the free fall of bodies through air—the behaviour of bodies of more complex geometry is of interest. When these autorotate, a lift force is generated which can significantly alter the body's trajectory. Here I examine the autorotation of prisms whose cross-sections are regular polygons. Prisms of triangular section rotate fastest, but generate less lift than a flat plate. Only bodies with less than eight sides are found to show autorotational behaviour. In all of these cases, the lift forces generated are larger than those obtained from a spinning cylinder driven externally at the same rotation speed. Many devices have been proposed in the past that use as a propulsion mechanism the lift force developed on a driven rotating cylinder, perhaps the most spectacular being the Flettner rotorship which crossed the Atlantic in 1926². The use of polygonal bodies would have application in these cases, not only because of the higher lift generated but also because the energy required may be derived simply from the relative wind. There are thus also clear implications for wind-power devices.

Whereas some bodies (such as windmill arms and samara seeds) held in a flowing stream will experience a torque causing them to rotate, there is a class of bodies that instead will take up a stable position and will not rotate unless given a sufficiently large initial impulse. As an example, a flat plate pivoted on a line of symmetry will take up an equilibrium position with its face normal to the airstream. If perturbed gradually from this position it will tend to return because of the net torque (resulting from the surface pressure distribution) acting to counter the rotation, as indicated in inset *a* of Fig. 1. At low speeds a net

positive torque is thus required to sustain rotation. But when a piece of stiff cardboard is dropped with an edge facing downwards it will start rotating about an axis normal to the direction of motion. In this case the rotation results in a strong vortex being developed behind the retreating face³. The resulting low pressure behind this face shifts the aerodynamic force so that it acts to assist rotation as indicated in inset *b* of Fig. 1. The body will thus accelerate until the aerodynamic moment balances the moment due to flow resistance, and stable autorotation at point A will result. At this condition the energy extracted from the air exactly balances that dissipated by the system. The equilibrium autorotational speed is a function of the shape of the body and the velocity of the airstream. The rotation generates a lift force (the Magnus⁴ effect). The balance between the lift, drag and weight results in the card following an angled trajectory to the ground.

The phenomenon of autorotation was initially reported by Maxwell⁵ more than 140 years ago. More recently there have been a number of quantitative studies on the autorotation of flat plates rotating about an axis perpendicular to the flow. Iverson⁶ correlated the results of a large number of experimental studies, both from wind-tunnel and free-flight tests. He showed that above a certain value of moment of inertia, the rotational speed becomes independent of moment of inertia, and depends only on the thickness-to-width ratio of the cross-section normal to the axis of rotation, and the aspect ratio (ratio of the length of the plate, along the axis, to the width). In my tests the limiting moment of inertia was exceeded by a significant margin.

The existence of the vortices (which are shed twice per revolution) is graphically shown in the work of Lugt³. He solved the governing Navier–Stokes equations numerically for the case of a two-dimensional (infinite aspect ratio) cylinder of elliptical section with a thickness ratio of 0.1. His results predicted stable autorotation (point A in Fig. 1) at a tip-speed ratio (ratio of the tangential velocity of the edge of the plate to the velocity of the oncoming airstream) of 0.45.

The above studies concentrated on the autorotational characteristics of flat plates, where the ratio of thickness to width was ~ 0.1 . In a recent experimental study⁷ simulating the infinite-aspect-ratio situation, I extended the data to cover all possible thickness ratios, and obtained excellent agreement with the value calculated by Lugt³. An unexpected result from this study was that the tip-speed ratio for the thin flat plate was the same as that for a body of square section, notwithstanding that four vortices are shed per revolution for the four-sided body, and only two for the flat plate. This raised the question of the autorotational characteristics of bodies with different numbers of sides. As it is known that circular cylinders do not autorotate, it was also of interest to determine whether there is a limiting number of sides for autorotation to occur.

The apparatus and methodology used here are the same as those described previously⁷. The model was mounted on low-friction bearings in a wind tunnel, spanning the full width of the tunnel. It was attached to a two-component force balance enabling the lift and drag forces to be measured. Prisms of

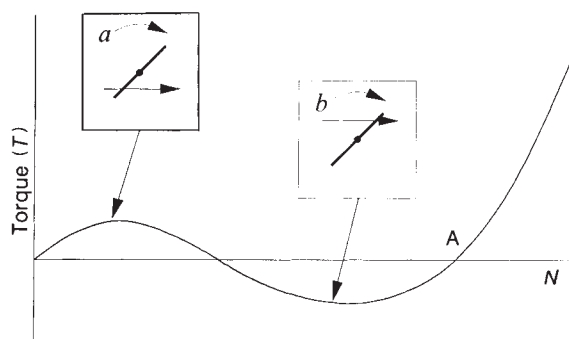


FIG. 1 External torque (T) required to drive the rotation of a body in an airstream, as a function of rotation speed (N)⁹. At *a*, the body will tend to return to a stationary position unless a torque is applied. At *b*, a torque is required to prevent rotation; otherwise the body will accelerate to point A, at which it autorotates.

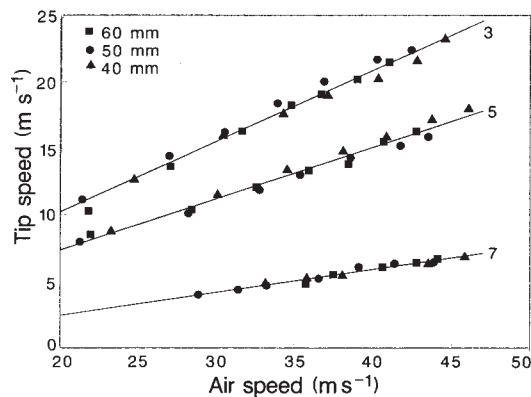


FIG. 2 Tip speed as a function of airstream speed for three-, five- and seven-sided bodies. In each case the tip-speed ratio (tip speed: air speed) is constant and independent of diameter.

polygonal cross-section with three to eight sides were manufactured from a low-friction rigid plastic, with diameters of a circumscribing circle of 40, 50 and 60 mm. Thin plates with a thickness less than 10% of the width were taken as representing the behaviour of two-sided plates.

Results of the measurement of tip speed V ($V = \pi DN$, where D is the diameter of the swept volume and N is the number of rotations per second) for autorotating prisms are given in Fig. 2 for three-, five- and seven-sided polygons, as a function of the oncoming airstream velocity. The first point to note is that for a given number of sides the results are independent of diameter. Since the range of Reynolds number is limited, it is to be expected that the flows around the polygons are similar. This is consistent with the findings of Lugt and Iverson, but is considerably different from that of a rotating circular cylinder, where boundary layer effects give considerable sensitivity to Reynolds number for similar tip-speed ratios. Second, tip speed is proportional to air speed, a result which is to be expected from dimensional analysis. The fact that the regression curves shown do not exactly pass through the origin is a result of the finite frictional torque of the bearings. Iverson showed that this effect may be accounted for by taking the slope of the curve (dV/dU) as a measure of the tip-speed ratio. The correction for the current tests is small because of the low pivot friction.

The autorotational behaviour may thus be characterized by the tip-speed ratio. Figure 3 gives the results for all the models tested. Note that the triangular prism spins the fastest. It therefore seems that the two- and four-sided prisms rotate at the same nondimensional speed by coincidence, as they lie on either side of the three-sided case. Many attempts were made to get the eight-sided model to autorotate but without success.

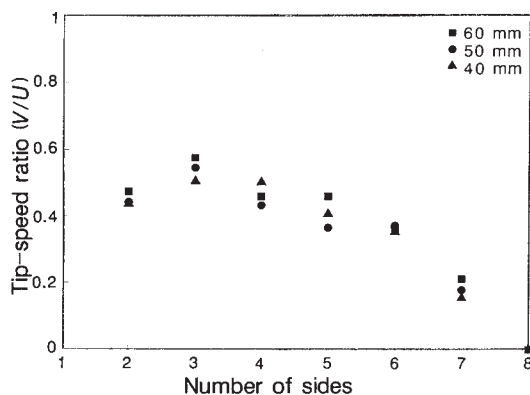


FIG. 3 Tip-speed ratio for bodies with cross-sections having up to eight sides. The highest tip-speed ratio is seen for the triangular prism. Despite repeated efforts, the eight-sided prism could not be made to autorotate.

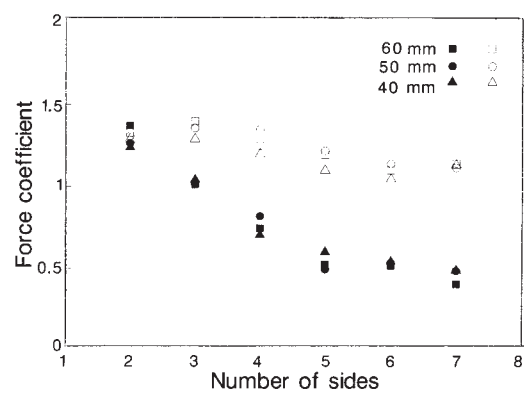


FIG. 4 Force coefficients for bodies with 2-8 sides. Open symbols indicate drag, filled symbols indicate lift. The lift decreases monotonically as the number of sides is increased, but remains higher than it would be for a circular cylinder driven at the corresponding tip-speed ratio.

The force coefficients are given in Fig. 4. The lift decreases monotonically with the number of sides. In all cases the lift coefficient is more than double the maximum obtained for a circular cylinder driven externally at the corresponding tip-speed ratio². Potential flow theory⁸ is a useful method for predicting lift forces on an aerofoil and uses the flow around a rotating cylinder as an intermediate step. It does not allow for the inclusion of viscosity, and thus boundary layer and separation effects are not accounted for. In this theory the lift is a direct function of the rotation, and for a spinning body of arbitrary cross-section the theory predicts a lift coefficient of $C_L = 2\pi V/U$, which is significantly higher than the experimental results presented here. This theory indicates a higher lift for the triangular section than for other sections because of the higher tip-speed ratio. It is clear, however, that the mechanisms for autorotation of bodies of prismatic section are fundamentally connected with the flow separation at the sharp edges and the consequent development of a shed vortex.

The drag is reduced slightly as the number of sides increases, and is of similar magnitude to that which would be obtained for a circular cylinder at similar Reynolds numbers and tip-speed ratios.

The effect of reduced aspect ratio would be to decrease both the tip-speed ratio and the lift. Iverson's results for thin plates show this feature very clearly. His data extend to a maximum aspect ratio of 4. Reducing this to a ratio of 0.5 results in a reduction in tip-speed ratio from ~ 0.65 to ~ 0.25 . This is due to the breakdown of the two-dimensional vortex pattern⁷ and the reduction of the pressure difference across the plate at the tips.

Of particular interest is the glide angle γ of a body falling while autorotating. This is given by $\cot \gamma = L/D$. For bodies of high aspect ratio, say of the order of 10 and above, the results should be similar to those found in these tests. The best lift-to-drag ratio obtained is for the two-sided plate and has a value of about unity. Such a body would thus have a glide angle of $\sim 45^\circ$ to the horizontal. For a vehicle breaking up in the air this could have a considerable effect on the distribution of the debris. Note that for bodies with from four to seven sides, the glide angle is nearly constant. □

Received 8 April; accepted 11 June 1991.

1. Lugt, H. J. *Ann. Rev. Fluid Mech.* **15**, 123-147 (1983).
2. Swanson, W. M. *J. basic Engng* **82**, 461-470 (1961).
3. Lugt, H. J. *J. Fluid Mech.* **99**, 817-840 (1980).
4. Magnus, G. *Poggendorfer Ann. Phys.* **88**, 604 (1853).
5. Maxwell, J. C. *Camb. Dublin Math. J.* **9**, 145 (1854).
6. Iverson, J. D. *J. Fluid Mech.* **92**, 327-348 (1979).
7. Skews, B. W. *J. Fluid Mech.* **217**, 33-40 (1990).
8. Kuethe, A. M. & Schetzler, J. D. *Foundations of Aerodynamics*, 2nd Edn (Wiley, New York, 1959).
9. Riabouchinsky, D. P. *J. Royal Aeronaut. Soc.* **39**, 282-348; 377-444 (1935).

Importance of *Phaeocystis* blooms in the high-latitude ocean carbon cycle

Walker O. Smith Jr*, Louis A. Codispoti†, David M. Nelson‡, Thomas Manley§, Edward J. Buskey||, H. Joseph Niebauer¶ & Glenn F. Cota#

* Botany Department and Graduate Program in Ecology, University of Tennessee, Knoxville, Tennessee 37996, USA

† Monterey Bay Aquarium Research Institute, Pacific Grove, California 93950, USA

‡ College of Oceanography, Oregon State University, Corvallis, Oregon 97331, USA

§ Department of Geology, Middlebury College, Middlebury, Vermont 05753, USA

|| Department of Oceanography, University of Texas, Port Aransas, Texas 78373, USA

¶ Institute of Marine Sciences, University of Alaska, Fairbanks, Alaska 99775, USA

Graduate Program in Ecology, University of Tennessee, Knoxville, Tennessee 37996, USA

THE Greenland Sea is particularly important to the world ocean circulation, and potentially to carbon dioxide exchange between the ocean and atmosphere, because it is an area of surface convergence and deep-water formation¹⁻³. Previous investigations indicate that biological productivity is low^{4,5} in this area, especially in waters remote from the ice edge. During April and early May 1989, however, we observed the development of a massive bloom of the colonial prymnesiophyte *Phaeocystis pouchetii* across much of the Greenland Sea. From measurements of the rate of removal of nitrate from surface waters, we calculate that the average regional new production was about 40 g C m^{-2} during the 35-day period of our observations. This rate of new production is approximately equal to that observed in other hyperproductive polar regions, such as the Bering Sea and the Bransfield Strait. Because *Phaeocystis* blooms seem to be frequent and widespread in polar oceans^{4,6}, our results suggest that the Greenland Sea may be a larger sink of atmospheric carbon dioxide than has been previously thought.

The Fram Strait region in the northern Greenland Sea is the only deep-water entrance to the Arctic Ocean. On the eastern side of the strait, relatively warm and saline water of Atlantic origin is transported towards the pole⁷. As these waters move north, they cool (increasing the solubility of carbon dioxide), sink beneath the Arctic surface waters and ultimately contribute to the formation of North Atlantic deep water, one of the dominant deep-water masses of the world ocean. In such a region, high biological productivity and subsequent vertical flux during periods of stratification could lead to large downward fluxes of carbon. Previous studies indicated, however, that only moderate rates of primary production^{4,5,8,9} contribute to the observed ocean-atmosphere CO_2 gradients that occur in this region³.

We obtained data during April and May 1989 as part of the Coordinated Eastern Arctic Experiment. Three transects were occupied at $78^\circ 30' \text{ N}$ across the eastern Greenland Sea from the coast of Spitsbergen into the ice (Fig. 1). The line was first sampled from 10–12 April 1989, and the final transect was completed from 15–17 May. Salinity, temperature, density, nutrients¹⁰, particulate organic carbon and nitrogen concentrations, phytoplankton biomass (as chlorophyll concentrations) and zooplankton biomass were routinely determined every 20 km using oceanographic techniques. All stations except the one closest to Spitsbergen were in waters more than 1,500 m deep.

Density distributions at the start of the study showed stratification and relatively shallow mixed layers (less than 15 m near the ice; Fig. 2a). Nitrate concentrations (Fig. 2b) and silicic acid concentrations (Fig. 2c) were high throughout the upper 150 m and similar to those in deep water^{11,12}, indicating that little biological removal had occurred. The concentrations of particular organic carbon were higher in the surface layer (Fig. 2d), but were only twice as high as the concentrations observed below the euphotic zone. These observations suggest that the spring phytoplankton bloom had not proceeded far, although conditions (adequate irradiance and nutrients, and moderate mixed-layer depths) were favourable for phytoplankton growth¹³.

By the end of the study, a massive bloom of *P. pouchetii* had occurred. The transect was characterized by strong nutrient depletion over wide areas. The stratification was greater, particularly in the eastern sector (Fig. 3a), and nitrate concentrations were diminished throughout, with values less than $1 \mu\text{M}$ at a number of stations (Fig. 3b). Silicic acid concentrations were virtually unchanged, except in the low-salinity waters of the West Spitzbergen current (Fig. 3c), and clearly indicated that the phytoplankton assemblage responsible for most of the nitrogen removal was not siliceous; low biogenic silica concentrations confirmed this (D.M.N., unpublished data). Particulate organic carbon concentrations were very high, with maximal levels over 500 mg m^{-3} (Fig. 3d). The bloom's magnitude was largely unaffected by zooplankton grazing, although active grazing did occur.

The amount of new production (that is, the photosynthetic production available for removal from the system¹⁴ and based on nitrate) can be conservatively estimated by comparing the nitrate concentrations from the first and last transects (Figs 2b and 3b) and making a number of assumptions. First, we assumed that the two transects are representative of the area and that differences in nitrate concentrations are primarily the result of biological removal rather than of physical advection. This assumption is reasonable because data from before the bloom indicate that initial nitrate concentrations similar to those we observed are widespread in the Norwegian and eastern Greenland Sea^{11,12}. In other words, we did not sample the same waters repeatedly, but we assume that initial nitrate concentrations in all waters were nearly identical. Secondly, we assumed that diffusive supply and loss of nitrate were negligible relative to

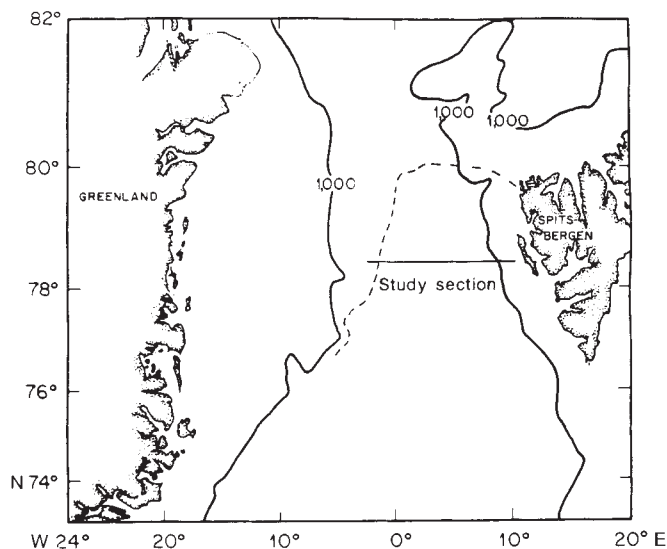


FIG. 1 Location of transects occupied across the Greenland Sea. The transects were occupied on 10–12 April and on 15–17 May 1989. Exact locations of stations sampled can be found elsewhere¹⁰. The dashed line marks the approximate position of the ice edge.

TABLE 1 Summary of rate information collected during Greenland Sea cruise

Parameter	Mean	Number of measurements	Standard deviation	Range
^{14}C Productivity ($\text{g C m}^{-2} \text{d}^{-1}$)	2.1	25	1.8	0.5–8.1
^{15}N nitrate uptake ($\text{mmol m}^{-2} \text{d}^{-1}$)	48	23	54	12–71
f -ratio	0.59	21	0.20	0.24–0.90

Data were integrated from the surface to the 0.1% isolume and generally included seven depths.

the rate of removal. Estimates of vertical diffusion of nitrate ($\partial\text{NO}_3^-/\partial z$) into the euphotic zone, obtained by multiplying an assumed vertical eddy diffusion coefficient ($5 \text{ cm}^2 \text{ s}^{-1}$) by the observed nitrate gradients ($\partial^2\text{NO}_3^-/\partial z^2$) at 100 m, suggest that neglecting this factor causes an underestimate of less than 5% in our estimate of new production. Finally, we assume that regenerative processes (particularly the production of nitrate by nitrification) are slow relative to removal processes¹⁰. This assumption should also favour a low estimate.

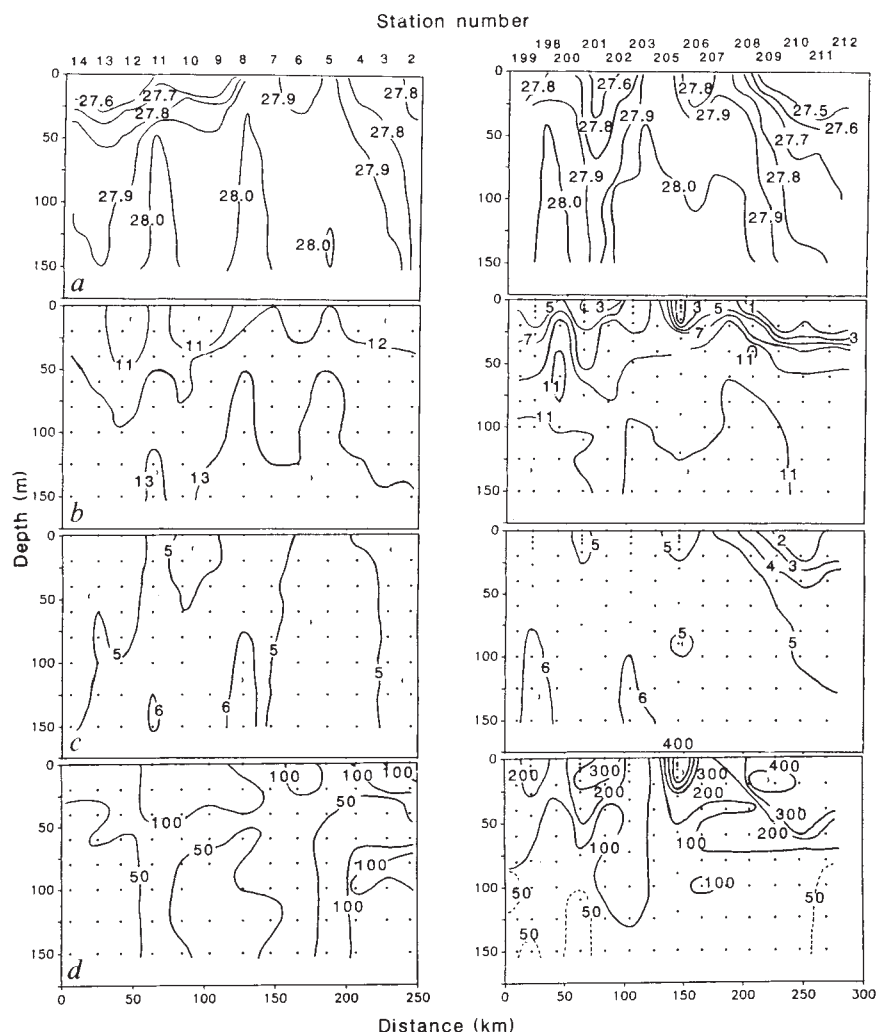
When we calculated the changes in nitrate concentrations over the 35-day period, we found that a large proportion had been removed. If the nitrate removal is integrated from the surface to a depth of 150 m and converted to carbon units using the Redfield ratio¹⁵, the new production is more than 40 g C m^{-2} over the 35-day period, corresponding to a mean daily new production of $1.2 \text{ g C m}^{-2} \text{ d}^{-1}$ over the entire section. The maximum within the section is more than $2.0 \text{ g C m}^{-2} \text{ d}^{-1}$. This is a substantial amount of new production over a very large

region. In addition, nitrate removal was apparent throughout the upper 150 m. We attribute this in part to the exceptional photoadaptive characteristics of the *Phaeocystis* populations¹⁶ as well as the relatively weak coupling between nitrogen uptake and irradiance in polar systems¹⁷. The mean of 25 measurements of ^{14}C primary productivity was $2.1 \text{ g C m}^{-2} \text{ d}^{-1}$; furthermore, nitrate uptake rates averaged $48 \text{ mmol m}^{-2} \text{ d}^{-1}$, and the mean f -ratio¹⁴ was 0.59 (Table 1).

All these measurements confirm the high primary productivity and elevated rates of nitrate-based production. Previous measurements of primary productivity have not suggested such high rates over so widespread an area. The mean productivity of the marginal ice zone (generally considered the most productive part of the region) in the summer of 1984 was $0.48 \text{ g C m}^{-2} \text{ d}^{-1}$, with a maximum rate of $1.8 \text{ g C m}^{-2} \text{ d}^{-1}$ (ref. 8), for example. The high productivity we observed was therefore surprising. Furthermore, the elevated productivity was not confined to the marginal ice zone but was present throughout the entire area that we sampled.

For *Phaeocystis* blooms to be important components of the polar carbon cycle, they must occur frequently. We cannot yet completely assess the occurrence and distribution of *Phaeocystis* because previous field observations have been limited. Blooms have, however, been reported in regions north of 75° in 1984⁴, 1985⁴, 1987¹⁸ and 1988 (J. Deming, personal communication), and moderate to large occurrences of *Phaeocystis* have been identified from an analysis of silicate anomalies in ten of the years since 1973 (56% of the years) in waters north of Iceland¹⁹. In 1989 it dominated throughout much of the Greenland Sea through the entire summer²⁰. Because of the limited number of

FIG. 2 The vertical distribution of *a*, density (expressed as σ_t), *b*, nitrate (μM), *c*, silicic acid (μM) and *d*, particulate organic carbon (mg m^{-3}) across a transect at $78^\circ 30' \text{ N}$ in the Greenland Sea. Left-hand panels are for 10–12 April, right-hand panels for 15–17 May 1989.



observations, we cannot adequately explain why the 1989 *Phaeocystis* bloom was so prevalent. We note that comprehensive biological observations from the Greenland Sea early in the growing season are rare and speculate that early blooms of *Phaeocystis* might be typical features.

The introduction of organic matter by means of vertical flux into the deep waters of the North Atlantic is an important process in the ocean CO₂ cycle. Because the productivity of the Greenland Sea was previously considered to be moderate, rates of vertical transport of organic matter have generally been considered to be low. Indeed, the few direct measurements of vertical flux to deep waters, at least for a few locations, confirm that the rates are generally low²¹. But the largest monthly vertical flux of organic carbon in a time series recorded from sediment traps at two locations in our study region occurred in May and were as much as 33 mg C m⁻² d⁻¹ (ref. 21), suggesting that spring blooms in this region contribute substantial amounts of organic matter to the deep sea (the locations were 78° 26.5' N, 4° 5.7' W and 78° 55.4' N, 6° 44.5' E; both near-bottom moorings in waters deeper than 1,100 m). We also found considerable pigment fluorescence at depth¹⁶. Recent studies have shown that *Phaeocystis* is ingested by the dominant macrozooplankton species in our area (*Calanus finmarchicus*, *C. glacialis* and *C. hyperboreus*)²², and that it rapidly sinks from the euphotic zone as aggregates⁶. Thus, accelerated flux of *Phaeocystis* could be mediated both by grazing and by particle aggregation.

It is worth comparing this new production with that of other polar regions. The new production of ice-edge phytoplankton blooms in the Southern Ocean has been estimated to be 50 g C m⁻² yr⁻¹ (ref. 23), and Karl *et al.*²⁴, using similar nutrient-depletion techniques, estimate the new production of a phytoplankton bloom in the Bransfield Strait (Antarctic peninsula) to be 1.7 g C m⁻² d⁻¹, similar to our observed rate. Finally, new production in the Bering Strait, one of the Arctic's most productive regions, calculated from nitrate uptake over a 15-day period, was 2.0 g C m⁻² d⁻¹ (ref. 25). It is clear that the new production in the Fram Strait during our experiment equalled that in regions that have been identified as exceptionally productive. Hyperproductive blooms of *Phaeocystis*, such as that reported here and the previously reported high productivities associated with dynamic features found later in the growing season²⁶, indicate that the world's most northerly open-ocean area may be much more productive than previously thought.

It has been suggested recently (from measurements of the difference in air-sea CO₂ concentrations) that the waters of the Northern Hemisphere are not as important sink for CO₂, and that the northern terrestrial biome is the main sink in the global carbon cycle²⁷. It is possible, however, that the air-sea differences were underestimated because of undersampling, and that this resulted in an underestimate of the quantitative role of northern oceans²⁸. This is particularly likely given the spatial and temporal variability of pCO₂ and of biological activity which has been observed in the North Atlantic²⁹ and in our study. If so, then the relative importance of the terrestrial and ocean regions of the Northern Hemisphere in the global carbon cycle cannot be determined until we have a more complete description of the variations in biological processes and air-sea CO₂ differences. □

10. Codispoti, L. A., Friedrich, G. E., Sakamoto, C. M. & Gordon, L. I. *J. mar. Systems* (in the press).
11. Jones, E. P., Nelson, D. M. & Treguer, P. in *Polar Oceanography, Part B* (ed. Smith, W. O. Jr) 407-476 (Academic, San Diego, 1990).
12. Clarke, R. A. *Deep-Sea Res.* **37**, 1383-1411 (1990).
13. Sverdrup, H. U. *J. Cons. Explor. Mer* **18**, 287 (1953).
14. Eppley, R. W. & Peterson, B. J. *Nature* **282**, 677-680 (1979).
15. Redfield, A. C. *Am. Scient.* **46**, 205-221 (1958).
16. Cota, G. F., Smith, W. O. Jr & Mitchell, B. G. *Eos* **71**, 79 (1990).
17. Smith, W. O. Jr & Harrison, W. G. *Deep-Sea Res.* (in the press).
18. Gunkel, J. thesis, Univ. of Kiel (1988).
19. Stefánsson, U. & Ólafsson, J. *J. Rit Fiskideld.* (in the press).
20. Bodungen, B. V., Bauerfeind, E., Koeve, W. & Zeitzschel, B. *Eos* **71**, 65-66 (1990).
21. Hebbeln, D. & Wefer, G. *Nature* **350**, 409-411 (1991).
22. Estep, K. W., Neijstgaard, J. C., Skjoldal, H. R. & Rey, F. *Mar. Ecol. Prog. Ser.* **67**, 235-249 (1990).
23. Smith, W. O. Jr & Nelson, D. M. *Limnol. Oceanogr.* **35**, 809-821 (1990).
24. Karl, D. M., Tilbrook, B. D. & Tien, G. *Deep-Sea Res.* (in the press).
25. Walsh, J. J. *et al. Prog. Oceanogr.* **22**, 277-359 (1989).
26. Smith, S. L., Smith, W. O. Jr, Codispoti, L. A. & Wilson, D. L. *J. mar. Res.* **43**, 693-707 (1985).
27. Tans, P. P., Fung, I. Y. & Takahashi, T. *Science* **247**, 1431-1438 (1990).
28. Keeling, C. D. *et al.* in *Aspects of Climate Variability in the Pacific and the Western Americas* (ed. Peterson, D.) 165-363 (American Geophysical Union, Washington DC, 1989).
29. Watson, A. J., Robinson, C., Robinson, J. E., Williams, P. J. le B. & Fasham, M. J. R. *Nature* **350**, 50-53 (1991).

ACKNOWLEDGEMENTS. We thank J. Deming and J. Grebmeier for criticisms of an early draft of the manuscript, and J. Sarmento for help with our discussion of the carbon cycle. This research was supported by the Office of Naval Research.

Volcanic production of polyphosphates and its relevance to prebiotic evolution

Y. Yamagata, H. Watanabe, M. Saitoh & T. Namba

Department of Physics, Faculty of Science, Kanazawa University,
1-1 Marunouchi, Kanazawa 920, Japan

PHOSPHATES would probably have been essential compounds for prebiotic evolution on the primitive Earth. In this context, there have been several studies of condensation of water-soluble phosphates to polyphosphates and phosphorylation and condensation or polymerization of biomolecules with polyphosphates¹⁻²⁰. But most of the phosphorus on the early Earth would have been in the form of water-insoluble apatite, and the origin of the water-soluble polyphosphates required for prebiotic evolution has therefore been a mystery²¹⁻²⁴. Here we show, both from experiments that simulate magmatic conditions and from analysis of volatile condensates in volcanic gas, that volcanic activity can produce water-soluble polyphosphates through partial hydrolysis of P₄O₁₀. This mechanism seems to be the only viable route identified so far for the production of these species on the primitive Earth.

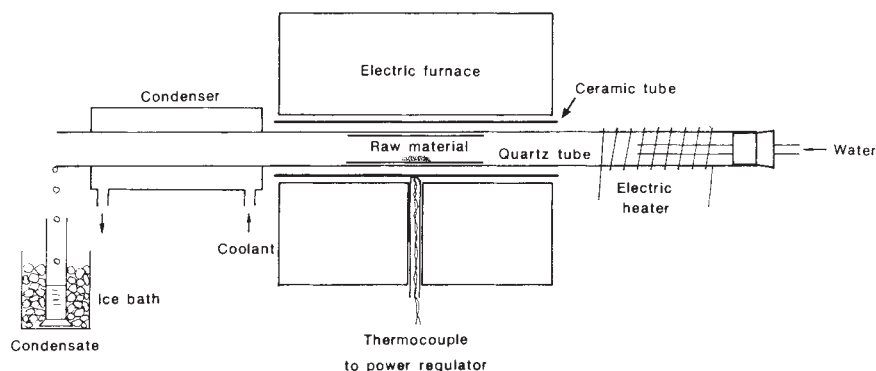
It has been suggested that the P₄O₁₀ molecule would be volatilized from magma^{25,26}. We have discussed¹⁹ another possible route for P₄O₁₀ formation, through the reoxidation by CO₂ of phosphorus volatilized from a reducing primitive Earth^{27,28}. As the primitive Earth might not have been strongly reducing, we have investigated the possibility of volatilization of P₄O₁₀ from a non-reducing magma at high temperature. The P₄O₁₀ molecule is thermodynamically stable¹⁹ even in the presence of water vapour at temperatures higher than 1,000 °C. If the volcanic gas containing P₄O₁₀ was suddenly cooled to near-normal temperature after the eruption, condensed phosphates might be produced by partial hydrolysis of the P₄O₁₀ molecules²⁹.

Our experiments were carried out using an apparatus that simulates sudden cooling and condensation of high-temperature volcanic gas, mainly composed of water volatilized from magma at high temperature (Fig. 1). As a model substance for the magma, finely powdered mixtures of basalt and Ca₃(PO₄)₂ or phosphate rock were used. Basalt powder was prepared from a basalt rock (Kagawa, Japan) by the standard method for geochemical analysis. The basalt analysis³⁰ showed it to contain: SiO₂ 51.52%; TiO₂ 0.94%; Al₂O₃ 18.60%; Fe₂O₃ 1.96%; FeO 6.44%; MnO 0.16%; MgO 6.22%; CaO 9.31%; Na₂O 2.64%; K₂O 1.22%; P₂O₅ 0.10%; H₂O(+) 0.75%; H₂O(-) 0.15%. The contents

Received 22 January; accepted 11 June 1991.

1. Peng, T.-H., Takahashi, T. & Broecker, W. S. *Tellus* **39**, 439-458 (1987).
2. Anderson, L. G. & Jones, E. P. *J. mar. Systems* (in the press).
3. Takahashi, T. *et al. J. Rit Fiskideld.* **9**, 20-36 (1985).
4. Baumann, M. E. M. thesis, Univ. of Aachen (1990).
5. Smith, W. O. Jr & Sakshaug, E. in *Polar Oceanography, Part B* (ed. Smith, W. O. Jr) 477-525 (Academic, San Diego, 1991).
6. Wassmann, P., Vernet, M., Mitchell, B. G. & Rey, F. *Mar. Ecol. Prog. Ser.* **66**, 183-195 (1990).
7. Paquette, R. G., Bourke, R. H., Newton, J. F. & Perdue, W. F. *J. geophys. Res.* **90**, 4866-4882 (1985).
8. Smith, W. O. Jr, Baumann, M. E. M., Wilson, D. L. & Aletsee, L. *J. geophys. Res.* **92**, 6777-6786 (1987).
9. Keene, N. K., Smith, W. O. Jr & Kattner, G. *Polar Biol.* (in the press).

FIG. 1 Experimental apparatus for simulating volcanic production of polyphosphates.



of carbon and nitrogen were less than 0.03 wt% and 0.01 wt% respectively, as measured by organic elemental analysis (Yanaco CHN Corder MT-3). The Kagawa basalt was used in most experiments, but another basalt (nephelinite, an unusual basalt abundant in alkali elements; P_2O_5 content, 2.09%) was also used for comparison.

The $Ca_3(PO_4)_2$ powder was purchased from the Wako Pure Chemical Co., and the phosphate rocks (from Morocco and South Africa, both granular) were given by the same company. The rock was ground to a fine powder in a ceramic mortar. A suspension of the powder (0.05 g) in 10 ml of 0.5 M H_2SO_4 was

ultrasonically vibrated and filtered, and the extract was analysed for phosphate by the molybdenum-blue method³¹. We found the content of P_2O_5 in the rocks to be 30.3% (Morocco) and 38.1% (South Africa). The basalts, $Ca_3(PO_4)_2$ and phosphate rock powder were ignited in air at 800 °C for more than ten hours to remove all organic material. The powdered basalt and phosphate samples were ground together in a ceramic mortar to produce a finely powdered mixture. The experiments were performed at temperatures of ~1,265 °C and ~1,340 °C at the centre of the furnace under which conditions the sample is molten. These temperatures are representative of those in

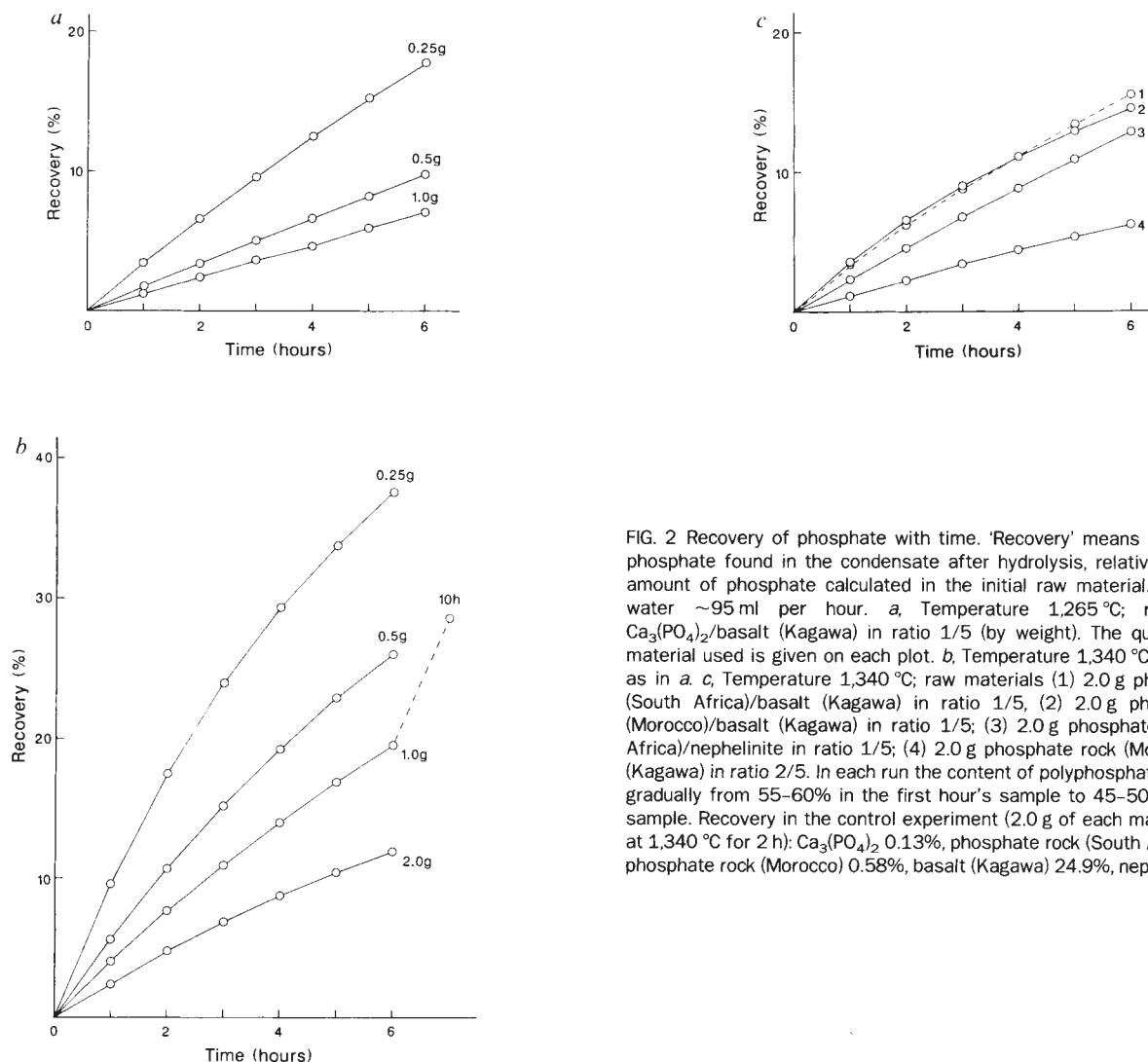


FIG. 2 Recovery of phosphate with time. 'Recovery' means percentage of phosphate found in the condensate after hydrolysis, relative to the total amount of phosphate calculated in the initial raw material. Flow rate of water ~95 ml per hour. *a*, Temperature 1,265 °C; raw material, $Ca_3(PO_4)_2$ /basalt (Kagawa) in ratio 1/5 (by weight). The quantity of raw material used is given on each plot. *b*, Temperature 1,340 °C; raw material as in *a*. *c*, Temperature 1,340 °C; raw materials (1) 2.0 g phosphate rock (South Africa)/basalt (Kagawa) in ratio 1/5; (2) 2.0 g phosphate rock (Morocco)/basalt (Kagawa) in ratio 1/5; (3) 2.0 g phosphate rock (South Africa)/nephelinite in ratio 1/5; (4) 2.0 g phosphate rock (Morocco)/basalt (Kagawa) in ratio 2/5. In each run the content of polyphosphates decreased gradually from 55–60% in the first hour's sample to 45–50% in the final sample. Recovery in the control experiment (2.0 g of each material, heated at 1,340 °C for 2 h): $Ca_3(PO_4)_2$ 0.13%, phosphate rock (South Africa) 0.17%, phosphate rock (Morocco) 0.58%, basalt (Kagawa) 24.9%, nephelinite 6.3%.

magmas³². To simulate the transport of volatile substances by volcanic gases (mainly water vapour) in the magma, water was supplied at one end of the tube at a rate of $\sim 95 \text{ ml h}^{-1}$, and the condensate sample forming in a condenser standing in a coolant at $\sim -2^\circ\text{C}$ was collected at the other end. We sampled the condensate at hourly intervals. The pH of the condensate was 3–4.5 in the first hour's sample and the pH increased gradually up to 5–6. Immediately after sampling, the sample collected in that hour was neutralized with sodium carbonate to avoid further hydrolysis of polyphosphates.

The phosphate in each condensate sample was analysed by the molybdenum-blue method, which allows the detection only of orthophosphate (PO_4^{3-})³¹. A direct analysis of the condensate gives us the amount of orthophosphate, and a second analysis after hydrolysis of polyphosphates (100°C , 1 hour in 0.5 M H_2SO_4) gives us the total amount of phosphate in the condensate. The quantity of phosphate present as polyphosphates is then estimated from the increment of phosphate after hydrolysis. Typical results of the experiments are shown in Fig. 2. As a control, both basalts, $\text{Ca}_3(\text{PO}_4)_2$ and the phosphate rocks were heated individually under the experimental conditions described above. The results are given in Fig. 2.

We analysed the condensates by ^{31}P NMR (JEOL Ltd, JNM GX-400, at 161.70 MHz, 10-mm-diameter sample tube). The condensate sample was first freeze-dried and dissolved in 4 ml of 50 mM EDTA. After adjusting the pH to 8.7 with 10 M NaOH and filtering (membrane filter: $0.45 \mu\text{m}$), the solution was analysed by NMR. EDTA was used to avoid broadening of the NMR lines by magnetic ions possibly present in the condensate. The pH value of 8.7 was chosen to avoid overlapping of NMR

lines. A typical spectrum is shown in Fig. 3a. The quantity of condensed phosphate as a percentage of the total phosphate indicated by the NMR analysis agreed well with that estimated by the molybdenum-blue method suggesting that contamination (by arsenic, for example) is not a problem.

Molecules of P_4O_{10} hydrolyse to orthophosphate and various polyphosphates in proportions that depend on the degree of hydrolysis²⁹. Figure 3b illustrates the production of polyphosphates by hydrolysis of phosphorus pentoxide under fairly severe conditions. The ^{31}P -NMR spectrum is very similar to that of Fig. 3a. Trimetaphosphate ($\text{P}_3\text{O}_9^{3-}$) and tetrametaphosphate ($\text{P}_4\text{O}_{12}^{4-}$), which are particularly useful^{10–12,19,20} in chemical evolution, survive under these severe conditions. Following an eruption, high-temperature and high-pressure volcanic gas (mainly consisting of water vapour) expands adiabatically in a very short time into the atmosphere. The gas will be spontaneously cooled to a lower temperature by the adiabatic expansion. If we assume that magmatic gas at $1,300^\circ\text{C}$ and 300 atm erupts into air at 0.5 atm and expands adiabatically, then a thermodynamic calculation indicates that the temperature of the gas might be decreased to 37°C . We can therefore expect a much higher yield of polyphosphates in volcanic clouds after the eruptions than in our experiments. The presence of water in the magma would enhance P_4O_{10} formation and volatilization because it markedly decreases the melting temperature and the viscosity³³.

Figure 2 shows that the percentage of phosphate recovered increases with decreasing quantities of raw material used in the experiment. This is a result of the change in ratio of surface area to volume of the melted raw material, because P_4O_{10} molecules evaporate from the surface. In reality, eruption of volcanos

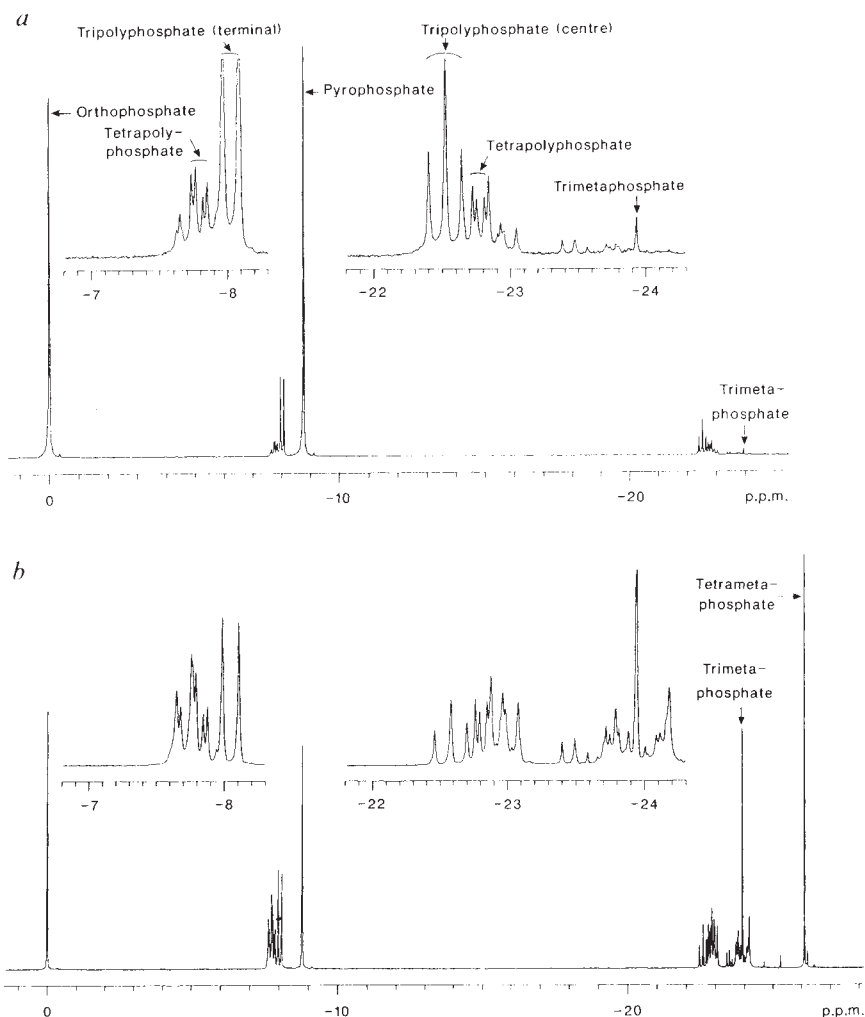
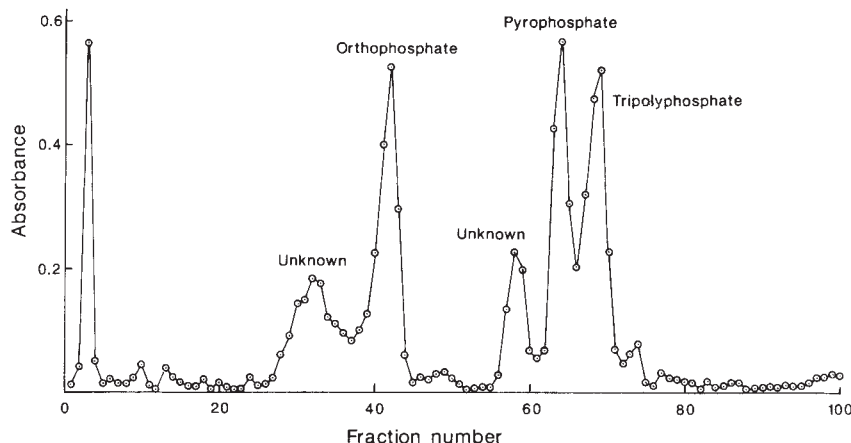


FIG. 3 a, Phosphorus-31 NMR spectrum (24,470 spectra accumulated) of condensate in the simulation experiment on sample (4) in Fig. 2c. The sample was prepared as described in the text and heated at 100°C for 10 minutes. After heating, the proportion of trimetaphosphate was several times as high as before heating, although the height of neighbouring peaks decreased considerably. The effect became more pronounced when heating was continued for a further 10 or 20 minutes. This suggests that the neighbouring peaks are due to substituted trimetaphosphate, ($\text{HP}_4\text{O}_{12}^{3-}$), which is relatively unstable and is hydrolysed to trimetaphosphate ($\text{P}_3\text{O}_9^{3-}$), tetrapolyphosphate ($\text{P}_4\text{O}_{13}^{3-}$) or isotetrapolyphosphate ($\text{H}_3\text{P}_4\text{O}_{13}^{3-}$)²⁹. Integrating the NMR peaks gave the following percentages: orthophosphate 39.4%, pyrophosphate 34.3%, tripolyphosphate 16.4%, tetrapolyphosphate 6.8%, trimetaphosphate 0.43%, unknown (probably contains substituted trimetaphosphate and isotetrapolyphosphate) 2.67%. b, Spectrum of aqueous solution of phosphorus pentoxide (90 mM, pH 8.7). The sample was prepared by dissolving the material in boiling water.

FIG. 4 Anion-exchange column chromatogram for the analysis of orthophosphate and polyphosphates in the volcanic gas condensate. Column: QAE-Sephadex A-25, 1.5×32.5 cm. Elution: 0.02–0.4 M NaHCO_3 (1,000 ml) linear gradient elution. Fraction: 10 ml, of which 8 ml was analysed by the molybdenum-blue method after heating to dryness followed by acid hydrolysis. The peak at the start of the chromatogram indicates the presence of some complex of polyphosphate with metal ions which cannot be eliminated by the cation-exchange column. Absorbance measured at 820 nm.



would result in bubbling in the interior of the magma because of the pressure decrease, and the area for the volatilization would be very large. We can therefore expect a high yield of P_4O_{10} in an eruption.

We have also obtained direct evidence for the production of polyphosphates in volcanic gas by collecting condensates of volcanic gas from several fumaroles at Mount Usu (Hokkaido, Japan) at temperatures of 540–690 °C. The samples were extracted by inserting one end of a quartz tube directly into the fumarole and cooling the other end. The condensates were immediately neutralized to avoid further degradation of the polyphosphates. We estimated concentration of orthophosphate in the samples after hydrolysis to be 6–9 μM , whereas before hydrolysis it was $\sim 1 \mu\text{M}$. These results indicate that a high proportion of polyphosphates was present. A portion (250 ml) of one of the condensates was condensed to 5 ml on a rotary evaporator. After filtering, the sample was analysed by ^{31}P NMR (number of spectra accumulated was 3.4×10^5). The spectrum showed peaks corresponding to ortho-, pyro- and tripolyphosphate and to an unknown phosphorus compound, although the peaks of pyrophosphate and the terminal phosphates of tripolyphosphate were not separated because of line broadening (probably by magnetic ions). The NMR sample was analysed by anion-exchange chromatography after passing it through a cation-exchange column to eliminate magnetic ions. The chromatogram is shown in Fig. 4. The analysis showed the presence of orthophosphate, pyrophosphate and tripolyphosphate (together with other unknown peaks) at concentrations estimated to be 1.13 μM , 0.45 μM and 0.37 μM , respectively. We have also analysed condensates of volcanic gas from fumaroles at low temperatures (104 °C and 94 °C), but no phosphates were found.

Here we propose a mechanism in which P_4O_{10} molecules are volatilized from the rock samples. All the phosphorus in the samples would be in the form of orthophosphate. The formation of P_4O_{10} therefore requires collision between phosphate ions, which can occur only in the melt, not in the solid phase. The basalt is molten (melting point $\sim 1,200$ °C) in our experiments and must have been a good solvent for $\text{Ca}_3(\text{PO}_4)_2$ (m.p. $\sim 1,670$ °C) and phosphate rock (m.p. $\sim 1,620$ °C). The P_4O_{10} molecules formed in the melt can then evaporate from the surface of the melt. This hypothesis explains the experimental results reasonably: $\text{Ca}_3(\text{PO}_4)_2$ or phosphate rocks alone do not produce significant quantities of phosphate, whereas basalts alone do; a mixture of basalt and $\text{Ca}_3(\text{PO}_4)_2$ does not produce soluble phosphate at a temperature ($\sim 1,100$ °C) lower than the melting temperature of the basalt; and the experiments at 1,340 °C give recoveries twice as high as that in the experiments at 1,265 °C, because of the greater mobility of phosphate groups at the higher temperature.

Recently, it has been proposed³⁴ that PO_2 (practically²⁹, this

would be in the form of $(\text{PO}_2)_n$) is the principal chemical form of phosphorus liberated from the Usu magma. Aqueous solutions of PO_2 are, however, difficult to oxidize to phosphoric acid²⁹. The two minor unknown peaks in Fig. 4 which are hard to attribute to hydrolysis products of P_4O_{10} probably originate from other sources, such as PO_2 .

It is commonly believed that the primitive atmosphere was formed by the vigorous release of gas from the interior of the Earth at a high temperature. This action would have permitted large-scale production of polyphosphates on the primitive Earth. Phosphate from hydrolysis of polyphosphates would precipitate to the sea bed as insoluble salts. If the precipitates are brought to a volcanic area, they again volatilize as P_4O_{10} gas. Thus, phosphate must have been continuously recycled by volcanic processes throughout the Earth's history. □

Received 14 February; accepted 30 May 1991.

1. Ponnamperna, C., Sagan, C. & Mariner, R. *Nature* **199**, 222–226 (1963).
2. Schwartz, A. & Fox, S. W. *Biochim. biophys. Acta* **87**, 694–696 (1964).
3. Ponnamperna, C. & Madk, R. *Science* **148**, 1221–1223 (1965).
4. Waenheldt, T. V. & Fox, S. W. *Biochim. biophys. Acta* **134**, 1–8 (1967).
5. Schwartz, A. & Fox, S. W. *Biochim. biophys. Acta* **134**, 9–16 (1967).
6. Rabinowitz, J., Chang, S. & Ponnamperna, C. *Nature* **218**, 442–443 (1968).
7. Schwartz, A. & Ponnamperna, C. *Nature* **218**, 443 (1968).
8. Lohrmann, R. & Orgel, L. E. *Science* **161**, 64–66 (1968).
9. Ferris, J. P. *Science* **161**, 53–54 (1968).
10. Rabinowitz, J., Flores, J., Krebsbach, R. & Rogers, G. *Nature* **224**, 795–796 (1969).
11. Rabinowitz, J. *Helv. chim. Acta* **53**, 1350–1355 (1970).
12. Chung, N., Lohrmann, R., Orgel, L. E. & Rabinowitz, J. *Tetrahedron* **27**, 1205–1210 (1971).
13. Lohrmann, R. & Orgel, L. E. *Science* **171**, 490–494 (1971).
14. Handschuh, G. J., Lohrmann, R. & Orgel, L. E. *J. molec. Evol.* **2**, 251–262 (1973).
15. Lohrmann, R. *J. molec. Evol.* **6**, 237–252 (1975).
16. Hulshof, J. & Ponnamperna, C. *Origins of Life* **7**, 197–224 (1976).
17. Shoffstall, A. M. *Origins of Life* **7**, 399–412 (1976).
18. Yamagata, Y., Matsukawa, T., Mohri, T. & Inomata, K. *Nature* **282**, 284–286 (1979).
19. Yamagata, Y., Kojima, H., Ejiri, K. & Inomata, K. *Origins of Life* **12**, 333–337 (1982).
20. Yamanaka, J., Inomata, K. & Yamagata, Y. *Origins of Life* **18**, 165–178 (1988).
21. Gulick, A. *Ann. N.Y. Acad. Sci.* **69**, 309–313 (1957).
22. Miller, S. L. & Urey, H. C. *Science* **130**, 245–251 (1959).
23. Miller, S. L. & Parris, M. *Nature* **204**, 1248–1250 (1964).
24. Schwartz, A. *Chemical Evolution and the Origin of Life, Vol. I Molecular Evolution* (eds Buvet, R. & Ponnamperna, C.) 207–215 (North-Holland, Amsterdam, 1971).
25. Griffith, E. J. *Chemical Evolution of the Early Precambrian* (ed. Ponnamperna, C.) 61–67 (Academic, New York, 1977).
26. Griffith, E. J. Ponnamperna, C. & Gabel, N. W. *Origins of Life* **8**, 71–85 (1977).
27. Ross, W. H., Mehrling, A. L. & Jones, R. M. *Ind. Engng Chem.* **16**, 563–566 (1924).
28. Schwartz, A. *Molecular Evolution: Prebiological and Biological* (eds Rohlfing, D. L. & Oparin, A. I.) 129–140 (Plenum, New York, 1972).
29. Van Wazer, L. R. *Phosphorus and its Compounds, Vol. I* (Interscience, New York, 1958).
30. Sato, H. *Sci. Rep. Kanazawa University* **26**, 51–72 (1981).
31. Chen, P. S., Toribara, T. Y. & Warner, H. *Analyt. Chem.* **28**, 1756–1758 (1956).
32. Tatsumi, Y., Sakuyama, M., Fukuyama, H. & Kushiro, I. *J. geophys. Res.* **88**, 5815–5825 (1983).
33. Best, M. G. *Igneous and Metamorphic Petrology* (Freeman, New York, 1982).
34. Manbo, V. S., Yoshida, M. & Matsuo, S. *J. Volcanol. geotherm. Res.* **46**, 37–47 (1991).

ACKNOWLEDGEMENTS. We thank S. Matsuo for donating the volcanic gas condensates, H. Satoh for the basalt powder, S. Arai for nephelineite, members of the Geology Department, Kanazawa University, for the use of their electric furnace, S. Matsuo, H. Satoh, M. Akaogi and K. Inomata for discussions, Y. Itatani and I. Okunaka for the organic elemental analysis, members of the inorganic chemistry group in Kanazawa University and Y. Kondo, K. Sasaki and K. Minagawa for their help, and J. Houghton and J. Paterson for reading the manuscript.

Fast rise times and the physical mechanism of deep earthquakes

H. Houston & Q. Williams

Institute of Tectonics, Earth Sciences Board, University of California, Santa Cruz, California 95064, USA

EARTHQUAKES at depths of >300 km are similar to shallower events in that they are dominantly of double-couple character¹, implying that shearing motion has taken place at depth. But because increased friction at these high pressures inhibits brittle fracture^{2,3}, various other mechanisms, related to phase transformations, have been invoked to explain the occurrence of deep earthquakes^{4–10}. As yet, however, no consistent differences have been found between the source characteristics of deep (>300 km) and intermediate-depth (<300 km) earthquakes^{2,11–13}. Here we report a systematic global survey of the rise times and stress drops of deep and intermediate earthquakes. (The rise time is defined as the time from rupture initiation to peak moment release rate.) When the rise times are scaled to the seismic moment release of the events, their average is nearly twice as fast for events deeper than ~ 450 km as for shallower events. This difference may ultimately provide an experimental means of testing proposed mechanisms for the generation of deep seismicity.

The notion that phase transformations may explain the existence of deep earthquakes is nearly half a century old¹⁴, but has recently gained support from laboratory experiments. The observation of high-pressure shear failure in ice and Mg_2GeO_4 undergoing metastable phase transitions^{4–6} led to the proposal that seismicity may be generated by the transformation of metastable $(\text{Mg, Fe})_2\text{SiO}_4$ to high-pressure phases in the subducting slab. More recently, faulting was observed¹⁵ in a natural olivine-bearing sample at pressures of 15 GPa (corresponding to 450 km depth in the Earth). This failure seems to be associated with the conversion of silicate olivine to β -phase under differential stress, generating a fine-grained, superplastic shear zone composed of the high-pressure phase¹⁵. Also, the amorphization of serpentine $(\text{MgFe})_3\text{Si}_2\text{O}_5(\text{OH})_4$, a hydrous phase thought to be present in the crustal portion of subducting oceanic lithosphere, has been proposed⁹ to generate the failure associated with deep earthquakes, based on acoustic emissions observed at high pressures and temperatures.

We aim here to determine whether the seismic signature of earthquakes presumably generated by brittle failure at intermediate depths⁵ (≤ 300 km) differs from that of deeper earthquakes possibly generated by phase transformation. We have therefore measured rise times, spectral amplitudes, seismic energies and stress drops of 68 events deeper than 120 km recorded teleseismically by the Global Digital Seismic Network (GDSN) between June 1982 and February 1987. Broadband displacement P waveforms were constructed by simultaneously deconvolving the instrument response from long-, intermediate- and short-period GDSN records at each station. For earthquakes deeper than 120 km, the surface-reflected phases pP and sP arrive long enough after the direct P wave not to contaminate this phase. The P-wave displacement waveforms are therefore essentially far-field source time functions convolved with the effect of attenuation. Such waveforms contain source information at periods of 1–60 s. The times from the onset of the P wave to the peak displacement were measured at all available stations and averaged to produce a rise time for each event. The rise times, as expected, are proportional to the cube root of the seismic moment of the earthquake¹⁶ (Fig. 1). To remove the effects of earthquake size from our data set, we therefore divide each rise time by the cube root of the seismic moment normalized by 1×10^{26} dyne cm. Seismic moments were taken from the Harvard Centroid Moment Tensor catalogue¹⁷.

Figure 2 shows the scaled rise times. For earthquakes at depths

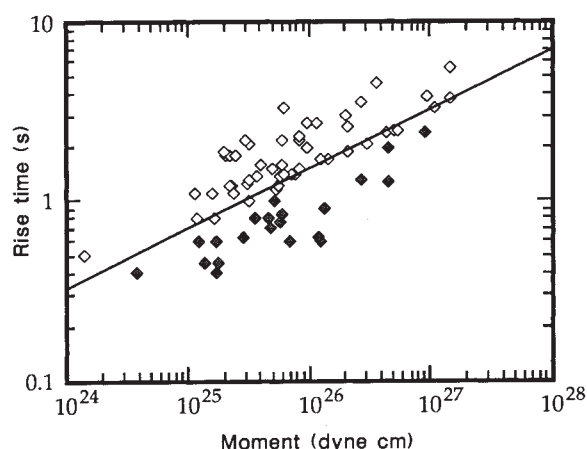


FIG. 1 Earthquake rise times against seismic moments. The line has a slope of $1/3$ showing the expected scaling relationship between rise times and the cube root of moment¹⁶ and is statistically indistinguishable from a least-squares fit to these data. Filled symbols indicate events with particularly rapid rise times for their moments.

of 120–400 km, the scaled rise times are greater than 1.4 s, whereas those at depths greater than 450 km have scaled rise times as short as 0.5 s. The maximum rise times observed at 450–700 km can be as long as those at shallower depths, but they show considerably larger variation. If the source is temporally complex, with multiple subevents spread over a considerable time, then this may lengthen the observed rise times; thus the minimum rise time at a given depth may be a more accurate indicator of the physical processes of rupture initiation. For about half of the events studied multiple subevents are clearly seen in the broadband records, consistent with previous results^{2,18}. In a few cases in which these subevents were readily distinguishable, we measured rise times for more than one subevent pulse, scaling the rise times with an appropriate moment for each subevent. No simple correlation was found between source complexity and rise time, but quantification of the amount of source complexity is difficult.

It is evident from Fig. 2 that average rise times decrease with depth. Although there are indications of trends in the scaled

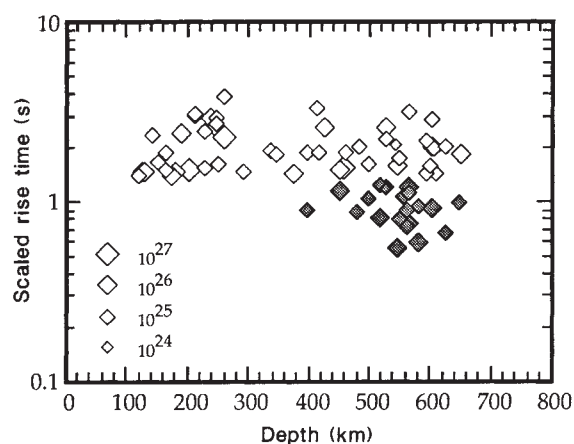


FIG. 2 Moment-scaled rise times against earthquake depths. The rise times have been scaled by the cube root of seismic moment to remove the effect of earthquake size (see text). The moment of each event is indicated by the size of the symbol (see key; units are dyne cm). Filled symbols indicate events with particularly rapid scaled rise times as in Fig. 1. Depths are those determined by the International Seismic Centre. About 45% of the events occurred in the Tonga–Kermadec subduction zone; their distribution of scaled rise times with depth mirrors that of the complete set.

rise times at depths of 100–300 km, the most robust feature of our data is the rapid onset for many events at depths greater than 450 km, giving events at 450–650 km an average rise time $\sim 60\%$ of that for shallower earthquakes.

One possible cause for the decrease in rise times with depth is the effect of attenuation, which tends to lengthen the rise times and durations of displacement pulses more for shallower events. This is simply because teleseismically recorded down-going waves from shallower earthquakes travel further through the relatively highly attenuating upper mantle than waves from deeper earthquakes. To evaluate this effect, we convolved a synthetic displacement pulse with a suite of Futterman Q -operators representing attenuation at different depths. Although depth-dependent attenuation has a significant effect on the observed duration of the source pulse, the effect on rise times is small, producing a difference of less than 0.2 s between rise times at depths of 250 and 600 km. Although the effect of structure in the slab is difficult to evaluate, such complexity would again affect observed source durations more than rise times. Furthermore, the probable effect would be to lengthen the observed rise times because of wavefront diffraction for the few seismic ray paths that traverse complex regions¹⁹.

According to simple kinematic fault models¹⁶, decreases in rise time may be caused by an increase in either the velocity of propagation of the rupture front or the particle slip velocity across the two sides of the shear zone, and we are at present unable to distinguish between these alternatives. To examine the dependence of stress drop on depth, we determined seismically radiated energies and stress drops. We integrated the square of the broadband velocity source spectrum to obtain seismic energy, and then applied the Orowan relation between seismic energy and moment²⁰ to obtain stress drops (Fig. 3a), using elastic parameters from PREM²¹. The stress drops, derived from high-quality broadband digital data, show no resolvable trend with depth, in accord with previous studies^{2,12,13}. This indicates that the faster initiation of ruptures below 450 km is not simply associated with an increase in the stress release in these events. The strain drop is proportional to the stress drop divided by the rigidity; Fig. 3b shows strain drops calculated from the stress drops by assuming a circular rupture area¹⁶. The decrease in strain drop with increasing depth indicates that the assumption of constant strain drop commonly made for shallow earthquakes¹⁶ is not appropriate over the depth range of subducted slabs.

Several mineralogical mechanisms could produce the change in rise times. Amongst the possible phase transformations in metastable olivine, the transformation of α -olivine to β -phase has now been observed to be associated with high-pressure faulting in a silicate assemblage¹⁵, the $\alpha \rightarrow \gamma$ -spinel transforma-

tion generates faulting in magnesium germanate^{4,5}, and metastable growth of small amounts of silicate γ -spinel has been observed in the stability field of β -phase²². A martensitic transformation from β -phase to γ -spinel has been proposed as a mechanism for some deep-focus events²². Both martensitic and incoherent nucleation-and-growth mechanisms may operate at depth, but we believe that the simplest explanation for faster rise times with depth may lie in the differing properties of the $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$ transitions. As β -(Mg, Fe)₂SiO₄ is the stable phase in the slab at depths of ~ 350 –450 km (ref. 23), the shorter rise times near the depth at which γ -spinel becomes stable may be caused by $\alpha \rightarrow \gamma$ transitions generating events with shorter onset times than $\alpha \rightarrow \beta$. First, release of latent heat is greater in the $\alpha \rightarrow \gamma$ transformation than in $\alpha \rightarrow \beta$ ²³; for a slab composed of 60% olivine, we calculate that a faulted zone undergoing the former reaction will be 25 K hotter than for the latter (presuming that the phase transition is synchronous with the faulting). Although the effect of a temperature rise on the rheology of a superplastic fault zone is poorly constrained, an increase in temperature would produce a less viscous fault zone, which in turn could allow faster slip velocities and hence faster rise times. The larger latent heat release from the $\alpha \rightarrow \gamma$ transformation may also more readily induce a thermal instability that could drive increased transformation and faulting. Second, the $\alpha \rightarrow \gamma$ transformation may proceed martensitically^{24,25}, generating more rapid shear than $\alpha \rightarrow \beta$ (although laboratory experiments have not found a correlation between regions of martensitic transformation and transformation-induced faulting^{4,5}). Alternatively, if the $\alpha \rightarrow \gamma$ transformation produces a smaller grain size than that generated by $\alpha \rightarrow \beta$, we would again expect a change in rheology of the faulted zone^{26,27}. But the relative grain sizes arising from these different transformations under slab conditions are not well constrained. Finally, we note that our observation of enhanced rise times at depths greater than 450 km is not simply explained by faulting associated with the amorphization of serpentine, as this process has been inferred⁹ to commence at 250 km.

We have suggested several mechanisms that could generate the shorter rise times observed below 450 km; the greater variability observed in these rise times may be related to increased heterogeneity at the source, produced by variable degrees of conversion of metastable olivine to its higher-pressure polymorphs. Increased heterogeneity in the slab has been proposed²⁸ as an explanation for the increase in aftershock activity observed below ~ 450 km.

The fast rise times observed for earthquakes deeper than 450 km suggest that they are generated by a different physical process than shallower events. Although we cannot yet differentiate between the possible mechanisms that could

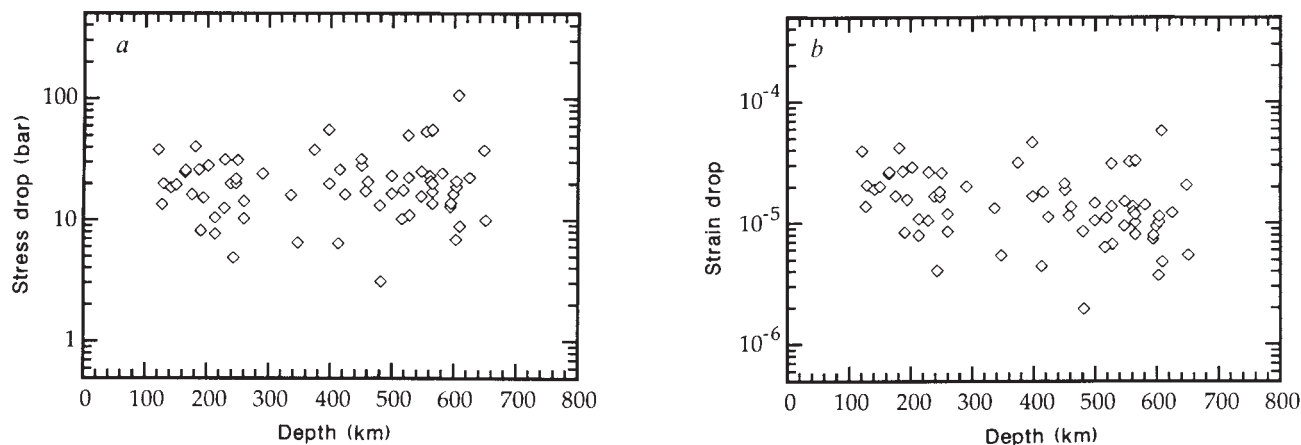


FIG. 3 a, Orowan stress drops against earthquake depths. The stress drops appear constant with depth. b, Strain drops against depths. The strain drops

are proportional to the stress drops divided by rigidity¹⁶, and hence decrease with increasing depth.

produce the shorter rise times from seismological arguments alone, our observations may allow proposed mechanisms of deep earthquake generation to be tested in the laboratory. With the advent of experimental techniques designed to measure acoustic emissions at high pressures⁹, scaled rise times may allow different types of failure induced by phase transformations to be evaluated for their consistency with the seismic signature of deep earthquakes. □

Received 6 February; accepted 19 June 1991.

1. Kawakatsu, H. *Nature* **351**, 50–53 (1991).
2. Frohlich, C. *Ann. Rev. Earth planet. Sci.* **17**, 227–254 (1989).
3. Griggs, D. T. & Handin, H. *Mem. geol. Soc. Am.* **79**, 347–373 (1960).
4. Green, H. W. & Burnley, P. C. *Nature* **341**, 733–737 (1989).
5. Burnley, P. C., Green, H. W. & Prior, D. J. *J. geophys. Res.* **96**, 425–553 (1991).
6. Kirby, S. H. *J. geophys. Res.* **92**, 13789–13800 (1987).
7. Sung, C. M. & Burns, R. G. *Tectonophysics* **31**, 1–32 (1976).
8. Liu, L.-G. *Phys. Earth planet. Inter.* **32**, 226–240 (1983).
9. Meade, C. & Jeanloz, R. *Science* **252**, 68–72 (1991).
10. Kirby, S. H., Durham, W. B. & Stern, L. A. *Science* **252**, 216–225 (1991).
11. Kuge, K. & Kawakatsu, H. *Geophys. Res. Lett.* **17**, 227–230 (1990).
12. Abe, K. *J. Phys. Earth* **30**, 321–330 (1982).
13. Wyss, M. & Molnar, P. *Phys. Earth planet. Inter.* **6**, 279–292 (1972).
14. Bridgman, P. W. *Am. J. Sci.* **A243**, 90–97 (1945).
15. Green, H. W., Young, T. E., Walker, D. & Scholz, C. *Nature* **348**, 720–722 (1990).
16. Kanamori, H. & Anderson, D. L. *Bull. seism. Soc. Am.* **65**, 1073–1095 (1975).
17. Dziewonski, A. M. & Woodhouse, J. H. *J. geophys. Res.* **88**, 3247–3271 (1983).
18. Fukao, Y. & Kikuchi, M. *Tectonophysics* **144**, 249–269 (1987).
19. Vidale, J. E. *Geophys. Res. Lett.* **14**, 542–545 (1987).
20. Houston, H. *Geophys. Res. Lett.* **17**, 1021–1024 (1990).
21. Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **25**, 297–356 (1981).
22. Rubie, D. C. & Brearley, A. J. *Nature* **348**, 628–631 (1990).
23. Akaogi, M., Ito, E. & Navrotsky, A. *J. geophys. Res.* **94**, 15671–15686 (1989).
24. Poirier, J. P. in *Anelasticity of the Earth*, *Geodyn. Ser. Vol. 4*, 113–117 (American Geophysical Union, Washington DC, 1981).
25. Burnley, P. C. & Green, H. W. *Nature* **338**, 753–756 (1989).
26. Vaughan, P. J. & Coe, R. S. *J. geophys. Res.* **86**, 389–404 (1981).
27. Ashby, M. F. & Verrall, R. A. *Phil. Trans. R. Soc. Lond.* **A288**, 59–95 (1978).
28. Frohlich, C. *J. geophys. Res.* **92**, 13944–13956 (1987).

ACKNOWLEDGEMENTS. We thank C. Frohlich, H. Green, S. Kirby, T. Lay, J. Revenaugh, S. Schwartz and J. Vidale for helpful comments, and H. Green, S. Kirby and R. Jeanloz for preprints. This research was partially supported by the NSF (H.H.) and NASA (Q.W.), and by the W. M. Keck Foundation.

Effects of a change in the level of inbreeding on the genetic load

S. C. H. Barrett* & D. Charlesworth†

* Department of Botany, University of Toronto, 25 Willcocks Street, Toronto, Ontario M5S 3B2, Canada

† Department of Ecology and Evolution, University of Chicago, 5630 S. Ingleside Avenue, Chicago, Illinois 60637, USA

“THE effects of inbreeding may not be as noticeable in the first generation as the invigoration immediately apparent after crossing”¹. This statement, published in 1919, has received little attention, and has apparently never been tested empirically, although the reduction of the genetic load of populations by inbreeding is well known in theoretical terms^{2–5}. Because inbreeding increases homozygosity, and hence the effectiveness of selection against recessive or partially recessive detrimental alleles, changes in levels of inbreeding can lead to a reduction in the frequencies of such mutant alleles. This results in equilibration at higher population mean fitness⁶ and is referred to as ‘purging’ populations of their genetic load. Severe inbreeding can also reduce genetic load due to overdominant alleles, provided selection coefficients are not symmetrical at all loci, because alleles giving lower fitness will be reduced in frequency at equilibrium^{7,8}. With either fitness model, however, reduction in genetic load takes time, and the initial effect of an increase in inbreeding is reduced fitness due to homozygosity. There are few data relating to the extent to which fitness is reduced during inbreeding in a set of lines and to how long the reduction lasts before increasing again to the initial level, or higher. Inbreeding experiments involving sib mating in mice and *Drosophila*

*subobscura*¹⁰, and successive bottlenecks in house flies¹¹ have yielded some evidence consistent with the purging hypothesis. Here, we report results of an experiment demonstrating a prolonged time-course of recovery of mean fitness under self-fertilization of a naturally outcrossing plant, and also compare our results with expectations derived by computer calculations. Our results show that the genetic load present in an outcrossing population can be explained only with a high mutation rate to partially recessive deleterious alleles, and that inbreeding purges the population of mutant alleles.

The experimental plant was the self-compatible annual water hyacinth, *Eichhornia paniculata* (Pontederiaceae). The species inhabits ephemeral pools and ditches in northeastern Brazil and the Caribbean, and exhibits a wide range of natural outcrossing rates¹². The experiments reported here used material from two populations with contrasting breeding systems. Population B11 from Brazil had an estimated outcrossing rate of 0.94, whereas

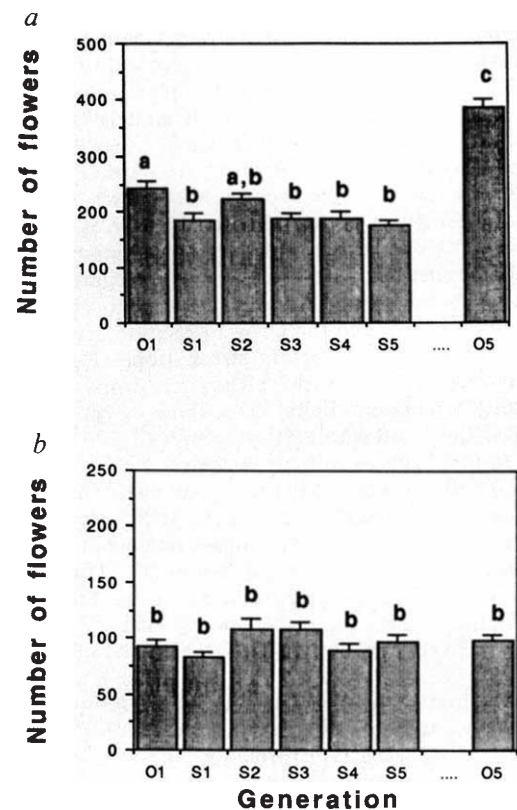


FIG. 1 Mean flower numbers of plants in various inbred and intercrossed generations. a, Naturally outbred population B11, NE Brazil. b, Naturally inbred population J13, Jamaica.

METHODS. Seed from open-pollinated families were randomly sampled from the two populations, B11 and J13. Lines were established in the glasshouse from each of 30 families from B11, and from each of 10 families from J13. One plant from each line was selfed and randomly outcrossed to other plants from the same population, to give the S_1 and O_1 generations, respectively. A small amount of S_1 seed was sown and a single randomly chosen plant was self-pollinated to give the S_2 generation. The remaining S_1 and O_1 seeds were stored dry at room temperature. The selfing was repeated for four successive generations with seeds from the S generations being stored. Plants from the lines were then selfed and outcrossed (within populations) to give the S_5 and O_5 generations. Stored seeds (maximum age 3 years) from all generations were then sown and the multi-generation plants grown in the glasshouse under uniform conditions. Fitness components were compared among generations in a randomized block design with all lines represented. Sample sizes were $n=1007$ for population B11 (average 4.8 plants per line per generation), and $n=378$ for J13 (average 5.4 plants per line per generation). Means and standard errors are shown, and the letters above the bars indicate statistically significant differences (5% level).

TABLE 1 Differences in mean fitness between different generations in our inbreeding experiments and models, and in experiments with maize¹⁷ and mice⁹.

Source of data	Inbred ₁	Inbred ₅	O ₅
<i>E. paniculata</i> (flower number)			
Population B11	0.761	0.732	1.597
Population J13	0.901	1.053	1.062
Maize (yield)			
1909	—	0.447	1.072
1910	—	0.394	1.124
1911	—	0.293	1.393
Mice (litter size)			
F ₁	—	0.924	1.021
F ₂	—	0.923	1.155
Theoretical			
Mutational model, initially completely outcrossing			
$s=0.2, h=0.2, U=1.0$	0.482	0.631	1.436
$s=0.2, h=0.1, U=1.0$	0.147	0.191	1.433
$s=0.2, h=0.2, U=0.5$	0.746	0.835	1.187
$s=0.9, h=0.2, U=0.1$	0.930	1.038	1.066
Mutational model, initial outcrossing rate of 0.9			
$s=0.2, h=0.2, U=1.0$	0.559	0.655	1.278
Overdominance model			
4 loci, $s_1=0.1, s_2=0.2$	0.873	0.782	0.994

Outcrossed generations are denoted by O, and inbred generations by "Inbred", and the numbers indicate the generation number when the fitness was estimated (see legend to Fig. 1). The fitness values are all expressed relative to the values for the initial outcrossed population (O₁), but for the maize and mouse data, values from families that were maintained by enforced cross-breeding are used. The numbers of generations of inbreeding for the maize experiment are not given in ref. 17, but there were certainly several generations of inbreeding. For the mouse experiment, the inbreeding coefficient was also non-zero (at least 0.3) when the experiment was stated⁹, and the data come from 11 generations of sib-mating followed by intercrossing to form F₁ and F₂ generations. The methods used to generate the theoretical results are given in the legend to Fig. 2.

population J13 from Jamaica was highly selfing, with no within-population heterozygosity observed at 21 allozyme loci¹³. A range of life-history traits was measured on all plants during the 8-month experiment. Only data on mean flower number per plant are presented here, but results for biomass were similar. Flower number exhibits the strongest inbreeding depression in *E. paniculata*. It represents the outcome of many developmental steps and must be influenced by many loci. As the species is annual, this trait is likely to be highly correlated with fitness. The results of five generations of self-fertilization of plants from population B11, followed by intercrossing the inbred lines, are shown in Fig. 1. Inbreeding had no significant effect on flower number per plant from the naturally inbreeding J13 population (Fig. 1; Table 1), showing that the differences observed are unlikely to be due to ageing of the stored seeds.

The expectations for such an experiment were calculated using deterministic computer programs that model either overdominant selection at several loci¹⁴, or mutation-selection balance at many unlinked loci⁵. There was remarkable similarity between the main features of the experimental results and the theoretical expectations of the mutational model for the inbred generations, when the initial populations before inbreeding begins were subject to a high mutation rate to moderately detrimental, partially recessive alleles (Figs. 1, 2; Table 1). Mean fitness values decreased in the first generation of inbreeding, but subsequently remained at roughly the same level. After five generations of inbreeding, the mean fitness of progeny derived from intercrossing the inbred lines was much higher than that of the inbred progeny, and also considerably higher than the fitness of the original outbred population. With strongly deleterious mutations (selection coefficient, 0.9), rapid recovery of fitness during the five generations of inbreeding was clearly evident (Fig. 2b; Table 1), unlike the experimental results. With

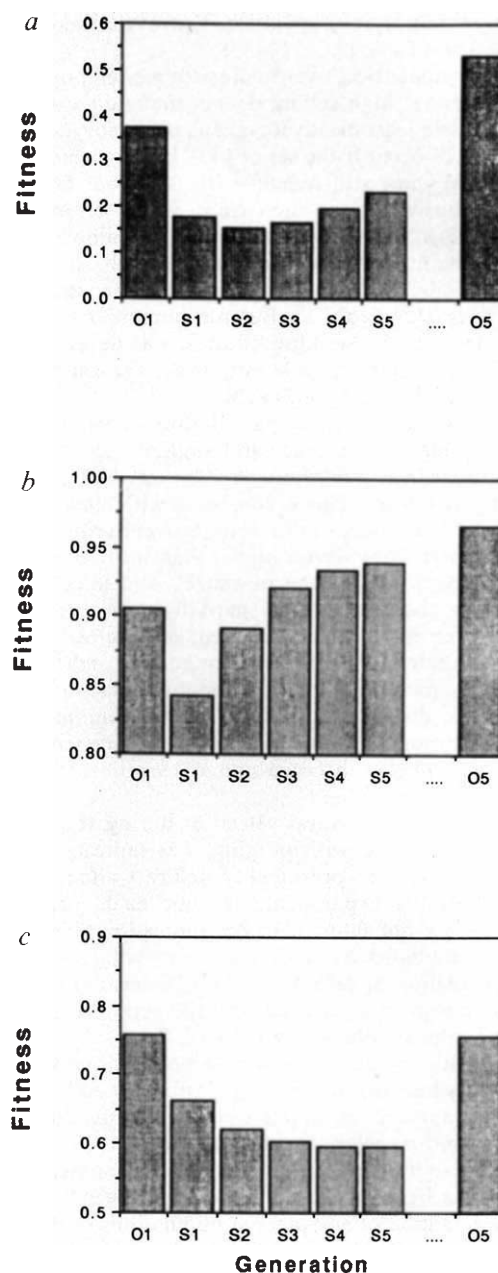


FIG. 2 Theoretical results for different selfed and intercrossed generations, assuming unlinked loci. a, Mutational model: selection coefficient $s=0.2$; dominance coefficient $h=0.2$; mutation rate per diploid genome $U=1.0$. b, Mutational model: $s=0.9$; $h=0.2$; $U=0.1$. c, Overdominance model: selection coefficients $s_1=0.1$ and $s_2=0.2$ at four loci.

METHODS. The computer calculations for the mutational model were obtained using the infinite population size model of Kondrashov²⁴ programmed as described⁶. In this model, the fitness of a homozygote for a mutant allele at one locus was denoted by $1-s$, where s is the selection coefficient, and the fitness of heterozygotes by $1-hs$, where h is the dominance coefficient. For the overdominance model, a five-locus deterministic program was used, with symmetric selection coefficients s_1 and s_2 (with the biologically implausible assumption of symmetrical overdominance, the mean fitness declines with inbreeding, even though variation is maintained with both alleles at frequencies of 0.5, and the fitness on intercrossing inbred lines will not exceed the original outbred value). To combine the fitness effects of different loci, multiplicative fitnesses were assumed in both models. For each model populations were run until equilibrium was reached under almost complete outcrossing (outcrossing rate 0.99), and then the selfing rate was altered to a high value (outcrossing rate 0.01). The time-course of changes in the population was then followed. Each generation, the fitnesses of inbred and outbred progeny produced in the population were calculated.

a partially inbred initial population, the results were similar to those observed (Table 1).

With the asymmetrical overdominance model, populations at equilibrium under high selfing do not maintain variation, but will fix the allele least disadvantageous in homozygotes though an increase will occur if the set of lines has not reached equilibrium, so that some still remain with the lower fitness allele. One would thus expect no increase in fitness on intercrossing the inbred lines after the population reaches equilibrium. During inbreeding, the fitness values behaved similarly to those in the mutational model, with a slight decrease, rather than increase, in fitness (Fig. 2c; Table 1). But the intercross mean fitness, although higher than the inbred fitness, was never higher than the fitness of the initial population, unlike the mutational load model and the observed data (Table 1).

These results support the partial dominance (mutational) hypothesis^{2,7} for genetic load, and suggest that the mutation rate per genome in *E. paniculata* must be very high (of the order of one per generation). This is consistent with most other available data^{15,16}. The failure of fitness to recover during inbreeding, and intercrossed fitness levels higher than those of the original outbred strains, are also seen in maize¹⁷ and mice⁹ (Table 1). This suggests that purging of partially recessive mutations accounts for an important component of heterosis. When the fitness values under inbreeding are plotted against the inbreeding coefficient, the partial dominance model yields an increase with high *F* values, due to purging, but the overdominance model predicts a monotonic decline. The partial dominance model can thus also explain the observed patterns in inbreeding experiments with conifers¹⁸.

The occurrence of natural selection during the inbreeding experiment, consistent with purging, was indicated by excess heterozygosity of five (presumably neutral) allozyme marker loci scored in the experimental plants, each generation of inbreeding (data not shown). Under inbreeding, heterozygotes tend to be produced by outcrosses. Surviving adults in the progeny generation therefore tend to be heterozygotes^{19,20}, and excess heterozygosity compared with the neutral expectation is frequently found in inbreeding plants²¹.

These results are also relevant to populations undergoing disturbance by humans. Inbreeding of formerly outbred populations occurs in many zoo populations, as well as domesticated populations and populations subjected to severe reductions in size as a result of human alterations of the environment. Extreme inbreeding has been recommended to purge genetic load and force the adaptation of endangered populations to the inbreeding regime they will experience under human management²². This assumes that lowered fitness inevitably caused by increased homozygosity during this process will not be too severe or prolonged. The validity of this assumption depends on the severity and dominance of the mutant alleles in the population being inbred^{23,5}. Our results indicate that purging by the most severe inbreeding, self-fertilization, could decrease fitness considerably, with little recovery under inbreeding, and that fitness is restored only when the inbred lines are intercrossed. □

Received 10 June; accepted 21 June 1991.

- East, E. M. & Jones, D. F. *Inbreeding and Outbreeding: Their Genetic and Sociological Significance* 147 (Lippincott, Philadelphia, 1919).
- Crow, J. F. *Genetics* **33**, 477–487 (1948).
- Wright, S. *Evolution and the Genetics of Populations* Vol. 3 (Univ. of Chicago Press, 1977).
- Lande, R. & Schmske, D. W. *Evolution* **39**, 24–40 (1985).
- Charlesworth, D., Morgan, M. T. & Charlesworth, B. *Evolution* **44**, 1469–1489 (1990).
- Crow, J. F. in *Mathematical Models in Population Genetics* (eds Kojima, K.-I.) 128–177 (Springer-Verlag, Berlin, 1970).
- Charlesworth, D. & Charlesworth, B. A. *Rev. ecol. Syst.* **18**, 237–268 (1987).
- Ziehe, M. & Roberts, J. H. *Genetics* **121**, 861–868 (1989).
- Barrett, S. C. H. & Husband, B. C. *Pl. Spec. Biol.* **5**, 41–55 (1990).
- Falconer, D. S. *Genet. Res. Camb.* **17**, 215–235 (1972).
- Hollingsworth, M. J. & Maynard-Smith, J. *J. Genet.* **53**, 295–314 (1955).
- Bryant, E. H., Meffert, L. M. & McCommas, S. A. *Amer. Nat.* **136**, 542–549 (1990).
- Barrett, S. C. H. & Husband, B. C. *Pl. Spec. Biol.* **5**, 41–55 (1990).
- Glover, D. E. & Barrett, S. C. H. *Heredity* **59**, 7–17 (1987).
- Charlesworth, D. & Charlesworth, B. *Evolution* **44**, 870–888 (1990).
- Kondrashov, A. *Nature* **336**, 435–440 (1988).

- Charlesworth, B., Charlesworth, D. & Morgan, M. T. *Nature* **347**, 380–382 (1990).
- Shull, G. H. in *Heterosis* (ed. Gowen, J. W.) 14–48 (Iowa University Press, Ames, 1950).
- Griffin, A. R. & Lindgren, D. *Theor. appl. Genet.* **71**, 334–343 (1985).
- Ohta, T. & Cockerham, C. C. *Genet. Res. Camb.* **23**, 191–200 (1974).
- Sing, C. F., Brewer, G. J. & Thirtle, B. *Genetics* **75**, 381–404 (1973).
- Brown, A. H. D. *Theor. appl. Genet.* **15**, 1–42 (1979).
- Templeton, A. R. & Read, B. *Zoo Biol.* **3**, 177–199 (1984).
- Wilton, A. N., Joseph, M. G. & Sved, J. *Genet. Res. Camb.* **54**, 129–140 (1989).
- Kondrashov, A. S. *Genetics* **111**, 635–653 (1985).

ACKNOWLEDGEMENTS. This research was funded in part by an operating grant from the Natural Sciences and Engineering Research Council of Canada (to S.C.H.B.) and in part by an NSF grant to D.C. We thank W. Cole for technical assistance and advice, and B. Charlesworth for his comments on the manuscript.

Transgenic plant aequorin reports the effects of touch and cold-shock and elicitors on cytoplasmic calcium

Marc R. Knight*, Anthony K. Campbell†, Steven M. Smith* & Anthony J. Trewavas*

* Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JH, UK

† Department of Medical Biochemistry, University of Wales College of Medicine, Cardiff CF4 4XN, UK

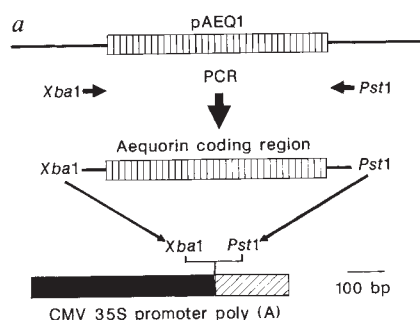
METHODS for measuring plant cytoplasmic calcium using micro-electrodes or microinjected fluorescent dyes are associated with extensive technical problems, so measurements have been limited to single or small groups of cells in tissue strips or protoplasts^{1,2}. Aequorin is a calcium-sensitive luminescent protein³ from the coelenterate *Aequorea victoria* (*A. forskalea*) which is formed from apoequorin, a polypeptide of relative molecular mass ~22,000, and coelenterazine, a hydrophobic luminophore⁴. Microinjected aequorin has been widely used for intracellular calcium measurement in animal cells⁴, but its use in plants has been limited to exceptionally large cells⁵. We show here that aequorin can be reconstituted in transformed plants and that it reports calcium changes induced by touch, cold-shock and fungal elicitors. Reconstituted aequorin is cytoplasmic and nonperturbing; measurements can be made on whole plants and a calcium indicator can be constituted in every viable cell. Now that apoequorin can be targeted to specific organelles, cells and tissues, with the range of coelenterazines with differing calcium sensitivities and properties available⁶, this new method could be valuable for determining the role of calcium in intracellular signalling processes in plants.

The apoequorin-coding region from complementary DNA clone pAEQ1 (ref. 7) was fused to the cauliflower mosaic virus (CMV) 35S promoter⁸ and transferred to *Nicotiana glauca* using the *Agrobacterium tumefaciens* pBIN19 binary vector system⁹ to provide constitutive expression (Fig. 1a). Expression was detected by western blots, using antibodies raised against aequorin purified from *Aequorea*¹⁰, in all plants transformed with the 35S-apoequorin chimaeric gene (Fig. 1b). F₁ progeny from the transformant expressing the highest levels of apoequorin (MAQ2.4; Fig. 1b) were selfed and homozygous F₂ progeny were used for subsequent experiments.

The time course for aequorin reconstitution from apoequorin in whole seedlings, in seedling homogenates and from apoequorin purified from *Aequorea*, all show similar kinetics (Fig. 2). Seedlings incubated in coelenterazine were transferred back to agar and continued to grow vigorously and respond tropically to light and gravity. No signs of toxicity were detected and growth rates of coelenterazine-treated and untreated seedlings were similar. The ratio of reconstituted aequorin in cotyledons, shoots and roots was 1:1:0.25 on a fresh-weight basis.

FIG. 1. Construction of chimaeric gene used to express apoaequorin in *Nicotiana*. *b*, Expression of apoaequorin in *Nicotiana*. Western blot analysis of total protein from the F₁ progeny of separate transformants. Lane 1, total protein from *E. coli* expressing apoaequorin¹⁹; lanes 2–7, separate *Nicotiana* 35S-apoaequorin transformants, MAQ3.3, MAQ3.2, MAQ2.13, MAQ2.4, MAQ2.3 and MAQ2.1; lane 8, wild-type untransformed *Nicotiana*.

METHODS. The apoaequorin coding region was amplified by polymerase chain reaction (PCR) from cDNA clone pAEQ1 (ref. 7) (the kind gift of Milton Cormier, University of Georgia), before construction of this chimaeric gene, to remove homopolymeric dC–dG tails present at either end of the cDNA insert. This counteracted inhibition of expression detected in constructs containing these tails (not shown). The *Xba*I–*Pst*I fragment was transferred to pDH51 (ref. 20) containing the CMV 35S promoter and terminator and the whole chimaeric gene was transferred to pBIN19 as an *Eco*R1 insert. All recombinant DNA manipulations and *Nicotiana* genetic transformations followed standard procedures^{21,22}. Total protein was prepared from 1 mg of 8-day transformant seedling tissue, electrophoresed on a 10% SDS-polyacrylamide gel and transferred to a nitrocellulose filter by electroblotting. After overnight incubation in a 1:1,000 dilution of mouse primary antibody



to aequorin isolated from *Aequorea*, the filter was incubated in a 1:5,000 dilution of alkaline phosphatase conjugated anti-mouse IgG for 2 h. Colour was developed by incubation in 100 μ g ml⁻¹ nitroblue tetrazolium, 60 μ g ml⁻¹ BCIP (5-bromo-4-chloro-3-indolyl phosphate), 4 mM MgCl₂, 100 mM Tris, pH 9.5. Protein electrophoresis and western blot analysis were according to standard procedures²³.

The cellular locale of the apoaequorin was estimated by centrifuging homogenates of transgenic seedlings at 116,000g and estimating distribution with coelenterazine by reconstitution. Over 97% of apoaequorin was contained in the soluble fraction.

To investigate the effect of touch, seedlings were placed in a luminometer and the cotyledons touched with a fine wire or plastic rod at intervals of 1 min (Fig. 3a). With each touch an intracellular calcium concentration ([Ca²⁺]_i) spike was observed. When repeatedly touched on the cotyledons at about 1-s intervals, a maximal response was elicited with about four touches (data not shown). Slight movement of the cotyledons on touching seems essential to this response. Touch responses are often thought to be limited to tendrils and sensitive plants such as *Mimosa pudica* and the Venus fly trap (*Dionaea muscipula*). Recent data¹¹, however, show that a variety of touch stimuli substantially reduces growth of *Arabidopsis* and other plants. In this case the involvement of [Ca²⁺]_i in transducing touch signals was surmised from increased calmodulin gene expression. The data in Fig. 3a support this hypothesis much more directly.

The effect of imposing temperature fluctuations on seedlings from an ambient 20°C to temperatures from 0°C to 50°C is shown in Fig. 3b. Dramatic increases in [Ca²⁺]_i accompany the transition to 0°C or 5°C. Such plants can be rewarmed and on recooling exhibit a further [Ca²⁺]_i spike. The spike is higher if 10 mM Ca²⁺ is included in the chilling medium but has no effect at higher temperatures. Only irrigation with Ca²⁺ at 100 mM has any effect at 20°C, causing repetitive spiking. No large increases due to heat-shock temperatures (40°C+) were observed. Many agronomically important plants are chill-sensitive and it has been suggested that chilling injury results from a failure in [Ca²⁺]_i homeostasis¹². This method of [Ca²⁺]_i measurement will permit a proper test of this hypothesis in chill-sensitive plants. Temperate plants use fluctuating temperature regimes as signals modifying dormancy, reproduction and inducing temperature adaptation. A role for [Ca²⁺]_i in transducing these signals is now clearly indicated.

A variety of fungal elicitors which initiate host-plant defence responses of hydrolytic enzyme production¹³, phytoalexin production and lignification¹⁴ have been reported in many species including tobacco¹⁵. Elicitor preparations from yeast and *Gliocladium deliquescens* have been shown to stimulate phytoalexin production in plant cells^{16,17}. These elicitor preparations also

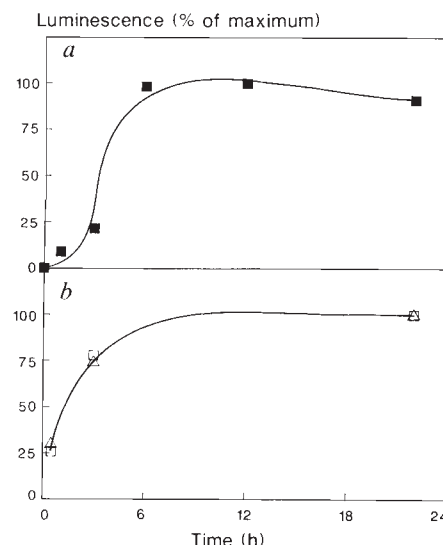


FIG. 2. Reconstitution of aequorin from apoaequorin in whole tobacco seedlings, tobacco homogenates and from purified *Aequorea* apoaequorin. *a*, Reconstitution time course of aequorin in whole *Nicotiana* seedlings (■). *b*, Reconstitution time course *in vitro* of apoaequorin isolated from *Aequorea* (△) and *Nicotiana* (□). Reconstitution was estimated by measuring the amount of calcium-dependent luminescence produced from the reconstituted aequorin at each time point.

METHODS. Apoaequorin was produced from *Aequorea* aequorin as previously described²⁴. Reconstitutions of apoaequorin in *Nicotiana* homogenates and from *Aequorea* were in 0.5 M NaCl, 5 mM mercaptoethanol, 5 mM EDTA, 0.1% gelatin (w/v), 10 mM Tris–HCl, pH 7.4, 2.5 μ M coelenterazine (gift from F. McCapra) at room temperature in darkness. Aequorin was reconstituted in whole *Nicotiana* seedlings by floating 8-day-old seedlings in water containing coelenterazine at 1 μ M in darkness. Before luminescence measurements, seedlings containing reconstituted aequorin were homogenized in 200 mM Tris, 0.5 mM EDTA, pH 7.0, and *in vitro* assays were diluted 1:50 in the same buffer. Reconstituted aequorin was then discharged by the addition of an equal volume of 50 mM CaCl₂ and the total amount of luminescence produced over 10 s measured. Reconstitution in whole *Nicotiana* seedlings was dependent on coelenterazine concentration up to 10 μ M. Reconstituted aequorin in tobacco was relatively stable, with more than 80% remaining after 48 h. To test the ability of reconstituted aequorin to report [Ca²⁺]_i, seedlings were treated with mild detergent and extracellular calcium ([Ca²⁺]_e) to different concentrations. This treatment produced increases in luminescence with increasing [Ca²⁺]_e (data not shown).

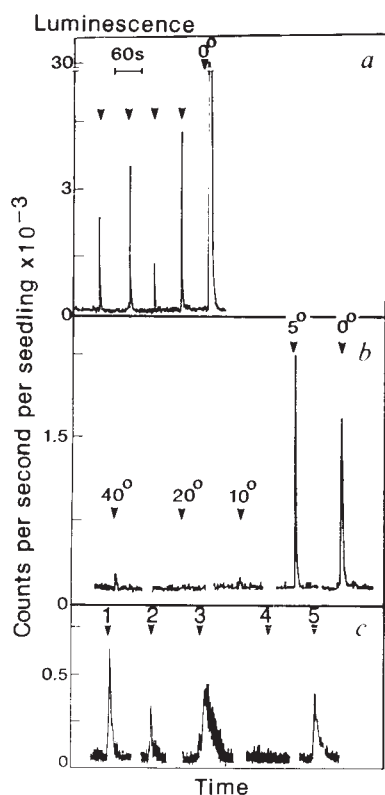


FIG. 3. The effect of touch-stimulus (a) temperature-shock (b) and fungal elicitors (c) on calcium-dependent aequorin luminescence.

METHODS. Transgenic tobacco seedlings (8 days old) were grown on half-strength M and S medium²⁵, 0.8% (w/v) agar. These were then incubated in water containing coelenterazine at 2 μ M (a, b) or 1 μ M (c) for 6 h in darkness. For luminescence measurements these seedlings were transferred singly to plastic cuvettes and placed in a luminometer. a, Seedlings were touched once every minute (at arrows) for 4 successive minutes with a fine wire, after which water at 0 °C was added. b, Seedlings were rapidly transferred from 20 °C to temperatures in the range 0–50 °C by the addition of water at the appropriate temperature. Only the treatments at 0 °C, 5 °C, 10 °C, 20 °C and 40 °C are shown. c, The following elicitor preparations were added: (1) untreated yeast elicitor¹⁶; (2) trifluoroacetic acid (TFA)-hydrolysed yeast elicitor; (3) void volume from a Sephadex G-25 column of yeast elicitor preparation; (4) proteinase K-digested yeast elicitor; and (5) untreated *Glucoladium deliquescens* elicitor¹⁷ (*G. deliquescens* spores a gift from E. Groskur). TFA hydrolysis was carried out at 120 °C for 1 h with TFA at 2 M. Proteinase K digestion was at 37 °C for 1 h at an enzyme concentration of 1 mg ml⁻¹. Luminescence was recorded by a digital luminometer and the output sent to a chart-recorder. All plants tested responded to the stimuli described. Increases in [Ca²⁺]_i showed low variability between plants in response to reproducible stimuli, for example, cold shock (b), but higher variability in response to unquantifiable stimuli such as touch (a). After measurements seedlings were removed from the luminometer and transferred back to agar, where both vigorous growth and tropic sensitivity were observed. Two other independent tobacco transformants (MAQ2.3 and MAQ2.13; Fig. 1b) showed the same responses (data not shown). No increase in luminescence was detected from untransformed seedlings treated with coelenterazine, nor from untreated transformed seedlings in response to the above stimuli (data not shown).

increase [Ca²⁺]_i (Fig. 3c). The active fraction in yeast is found in the void volume of a Sephadex-G25 column (Fig. 3c) but not G50 (data not shown); it shows some lability to trifluoroacetic acid under conditions that hydrolyse glycosidic linkages and is completely destroyed by proteinase K (Fig. 3c). Therefore, it is most probably a peptide of relative molecular mass (M_r) 5,000–30,000 (5–30K) with attached carbohydrate. Very few fungal elicitors have been characterized although some are carbohydrate and some protein¹⁴. Characterization of elicitors has so far required assays of enzymes induced in plant cells often over time courses of several days. This method of [Ca²⁺]_i measurement may greatly simplify elicitor purification and characterization owing to the rapidity of the assay. It has been speculated that elicitor signals are transduced through [Ca²⁺]_i (ref. 18). Again our data directly support this hypothesis.

The luminescence observed in these experiments (Fig. 3) can only be a qualitative measure of [Ca²⁺]_i variation. The numbers and types of cells responding cannot be determined and there will be physical barriers (for example, cell walls) to detection of luminescent light from cell layers deeper than two or three cells from the surface. Most probably epidermal cells are the

main contributor. The amount of aequorin discharged in any of these treatments is, however, below 5% of the total determined by homogenization and discharge. It cannot at present be determined whether this discharged aequorin can be regenerated. But the continuous *de novo* synthesis of apoequorin should allow reconstitution of aequorin several times in the same cell.

We are now analysing the source of the [Ca²⁺]_i increases described in Fig. 3 by preparing constructs that will target aequorin to plant organelles. The kinetics of the [Ca²⁺]_i transients induced by touch, cold shock and elicitors are all different. They are very short for touch, intermediate for cold-shock and long for elicitors. Different cellular sources for [Ca²⁺]_i may explain these variations. Without this considerable simplifying advance in technology these measurements would be very difficult to make. The availability of a range of coelenterazines with different calcium sensitivities⁶ can be expected to greatly improve measurements of different organelle calcium and provide additional flexibility to this technique. The goal of monitoring crop plant [Ca²⁺]_i in the field, which could be used diagnostically, is now brought much nearer with many consequent agricultural benefits. □

Received 24 May; accepted 2 July 1991.

- Gilroy, S., Hughes, W. A. & Trewavas, A. J. *FEBS Lett.* **199**, 217–221 (1986).
- Gilroy, S., Read, N. D. & Trewavas, A. J. *Nature* **346**, 769–771 (1990).
- Shimomura, O., Johnson, F. H. & Saiga, Y. *J. Cell. Physiol.* **59**, 223–240 (1962).
- Campbell, A. K. *Intracellular Calcium—its Universal Role as Regulator* (Wiley, Chichester, 1983).
- Williamson, R. E. & Ashley, C. C. *Nature* **296**, 647–651 (1982).
- Shimomura, O., Musicki, B. & Kishi, Y. *Biochem. J.* **261**, 913–920 (1989).
- Prasher, D. C., McCann, R. O. & Cormier, M. J. *Biochem. biophys. Res. Commun.* **126**, 1259–1268 (1985).
- Franck, G., Guille, M., Jonard, G., Richards, K. & Hirth, L. *Cell* **21**, 285–294 (1980).
- Bevan, M. *Nucleic Acids Res.* **12**, 8711–8721 (1984).
- Blinks, J. R. et al. *Meth. Enzym.* **57**, 292–328 (1978).
- Braam, J. & Davis, R. W. *Cell* **60**, 357–364 (1990).
- Minorsky, P. V. *Pl. Cell Envir.* **8**, 75–94 (1985).
- Mauch, F. & Staehelin, L. A. *Pl. Cell* **1**, 447–457 (1989).
- Dixon, R. A. *Biol. Rev.* **61**, 239–291 (1986).
- Watson, D. G., Rycroft, D. S., Freer, I. M. & Brooks, C. J. M. *Phytochemistry* **24**, 2195–2200 (1985).

- Hahn, M. G. & Albersheim, P. *Pl. Physiol.* **62**, 107–111 (1978).
- Brooks, C. J. W., Watson, D. G. & Freer, I. M. *Phytochemistry* **25**, 1089–1092 (1986).
- Ebel, J. & Grisebach, H. *Trends biochem. Sci.* **13**, 23–27 (1988).
- Prasher, D. C., McCann, R. O., Longiaru, M. & Cormier, M. in *Bioluminescence and Chemiluminescence—New Perspectives* (ed. Scholmerich, J.) 365–368 (Wiley, Chichester, 1986).
- Pietrzak, M., Shillito, R. D., Hohn, T. & Potrykus, I. *Nucleic Acids Res.* **14**, 5857–5868.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
- Draper, J., Scott, R., Armitage, P. & Walden, R. *Plant Genetic Transformation and Gene Expression—A Laboratory Manual* (Blackwell, Oxford, 1988).
- Harlow, E. & Lane, D. *Antibodies—A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1988).
- Campbell, A. K., Patel, A. K., Razavi, Z. S. & McCapra, F. *Biochem. J.* **252**, 143–149 (1988).
- Murashige, T. & Skoog, F. *Pl. Physiol.* **15**, 473–497 (1962).

ACKNOWLEDGEMENTS. We acknowledge financial support by the SERC and photographic assistance from W. Sinclair.

Normalization of current kinetics by interaction between the α_1 and β subunits of the skeletal muscle dihydropyridine-sensitive Ca^{2+} channel

Antonio E. Lacerda*, Haeyoung S. Kim†, Peter Ruth‡§, Edward Perez-Reyes*, Veit Flockerzi‡, Franz Hofmann‡§, Lutz Birnbaumer*†|| & Arthur M. Brown*

Departments of *Molecular Physiology & Biophysics and †Cell Biology and the ‡Division of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, USA
‡Institute of Physiological Chemistry, Faculty of Medicine, University of Saarland, D-6650 Homburg/Saar, Saarland, Germany
§Present address: Institute of Pharmacology, Technical University of Munich, D-8000 Munich, Germany

PURIFICATION of skeletal muscle dihydropyridine binding sites has enabled protein complexes to be isolated from which Ca^{2+} currents have been reconstituted. Complementary DNAs encoding the five subunits of the dihydropyridine receptor, α_1 , β , γ , α_2 and δ (ref. 1), have been cloned²⁻⁶ and it is now recognized that α_2 and δ are derived from a common precursor^{7,8}. The α_1 subunit can itself produce Ca^{2+} currents, as was demonstrated using mouse L cells lacking $\alpha_2\delta$ (refs 9, 10), β (ref. 10) and γ (our unpublished results). In L cells, stable expression of skeletal muscle α_1 alone was sufficient to generate voltage-sensitive, high-threshold L-type Ca^{2+} channel currents which were dihydropyridine-sensitive and blocked by Cd^{2+} , but the activation kinetics were about 100 times slower than expected for skeletal muscle Ca^{2+} channel currents. This could have been due to the cell type in which α_1 was being expressed or to the lack of a regulatory component particularly one of the subunits that copurifies with α_1 . We show here that coexpression of skeletal muscle β with skeletal muscle α_1 ,

generates cell lines expressing Ca^{2+} channel currents with normal activation kinetics as evidence for the participation of the dihydropyridine-receptor β subunits in the generation of skeletal muscle Ca^{2+} channel currents.

We previously described the construction of the LCa.11 cell line that resulted from transfecting pDHPR- α_1 into L(tk⁻) cells^{9,10}. pDHPR- α_1 is an expression vector based on pEV142 in which the dihydropyridine receptor (DHPR) α_1 subunit cDNA is under the control of a metallothionein promoter. These cells expressed DHP binding and Ca^{2+} currents but the binding and current density were low. To increase expression we prepared new cell lines by transfecting murine L(tk⁻) cells with pKNH-DHPR- α_1 as before⁹. The expression construct was engineered by subcloning the rabbit skeletal muscle DHPR α_1 subunit^{2,8} into the HindIII site of the pKNH expression vector, in which the inserted cDNA is under the control of the simian virus 40 (SV40) enhancer/promoter (for details, see ref. 11). G418-resistant cells, designated LCaN- α_1 cells, were selected and 49 clones screened initially for L-type Ca^{2+} channel currents in the presence of 0.5 μM Bay K8644.

Ca^{2+} channel currents in LCaN- α_1 cells were similar to those obtained in LCa.11 cells⁹. The currents were unaffected by changes in the holding potential between -50 and -90 mV, were stimulated by Bay K 8644, blocked by cadmium or cobalt and clearly belonged to the high-threshold or L category. Unfortunately the current densities remained low and were not different from those obtained with LCa.11 cells (Table 1). Thirty-three per cent (48/145) of the LCaN- α_1 cells from 49 clones having input resistances of about 10⁹ ohm gave rise to detectable tail currents in the presence of 0.5 μM Bay K8644; 52% (25/48) of these cells gave inward currents >5pA. This level was selected because it was more than twice the maximum inward current produced by errors associated with our linear correction procedure in control L cells. Just as for LCa.11 cells, the rate of current activation was very slow (Table 2), indicating that the slowness was not the result of induction of the metallothionein promoter with Zn^{2+} or the type of expression vector used previously. Activation time constants of Ca^{2+} channel currents in LCa.11 and LCaN- α_1 cells measured at test potentials of +30 or +40 mV were about 738 and 552 milliseconds, respectively

FIG. 1 *a*, Expression of α_1 mRNA in recipient LCaN- α_1 (clone 162) cells and of α_1 and β mRNAs in LCaN- α_1 derived cell lines after cotransfection with DHPR β in p91023(B) (p9-DHPR. β) and herpes simplex virus thymidine kinase gene in pHSV-106 (HSV-tk). The pKNH (ref. 11) and p91023(B) (ref. 12) vectors were gifts from S. Numa and R. J. Kaufman, respectively. Ten cell lines resistant to 400 $\mu\text{g ml}^{-1}$ G418 (selective pressure for conservation of α_1 DNA sequences) and able to grow in HAT medium (selection for expression of thymidine kinase and coexpression of β) were cloned, examined and tested for the presence of α_1 and β mRNA. Six out of the ten G418- and HAT-resistant cell lines expressed both DHPR components. Probes α_1 and β were labelled and blots were treated as described¹⁰. BC3H1 is a mouse cell line with a skeletal muscle phenotype. Migration positions of 28S and 18S ribosomal RNA markers are indicated. *b*, Specific DHP binding to membranes from cells expressing skeletal muscle DHPR α_1 before and after transfection with DHPR β . The cell lines were the same as analysed for α_1 and β mRNA in *a*. Top, specific binding at 0.32 nM [³H]-(+)-PN 200-110 to different LCaN- α_1 transfectants. Non-specific binding was determined with 10 μM unlabelled nitrendipine. Bottom, Scatchard analysis of DHP-binding sites in cell membranes of LCaN- α_1 clone 162 (K_d = 0.14 nM; open circles) and in cell membranes of one of its transfected DHPR β positive derivatives, LCaN- $\alpha_1\beta$ clone 4 (K_d = 0.23 nM; filled circles).

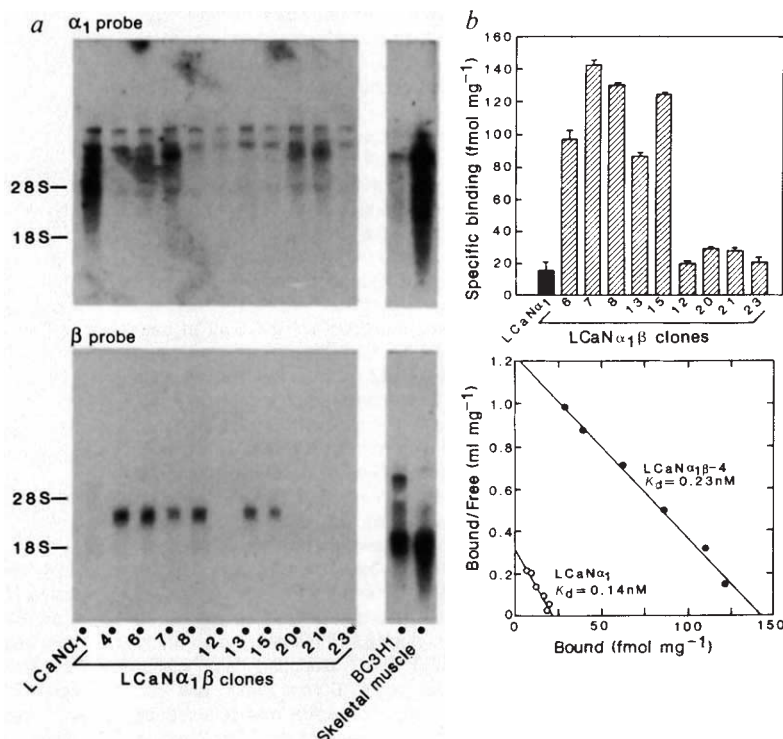


TABLE 1 DHP binding and current density characteristics of murine L cells transfected with the α_1 and β components of the rabbit skeletal muscle DHP receptor complex

Cell line	Component(s) expressed	Expression vector	Selection	DHP binding ([³ H]PN200-110)*		Current density†
				K_d (nM)	B_{max} (fmol mg ⁻¹)	
LCa.11	α_1	pEV-DHPR. α_1	HAT (tk ⁺)	0.38	61	0.73 ± 0.72 (22)
LCaN- α_1	α_1	pKNH-DHPR. α_1	G418	0.16	21	0.63 ± 0.45 (25)
LCaN- $\alpha_1\beta$ clone 13	$\alpha_1\beta$	+p9-DHPR. β	G418 plus HAT	0.17	145	0.74 ± 0.31 (7)
LCaN- $\alpha_1\beta$ clone 4	$\alpha_1\beta$	+p9-DHPR. β	G418 plus HAT	0.23	160	0.14 (1)
LCaN- $\alpha_1\beta$ clone 7	$\alpha_1\beta$	+p9-DHPR. β	G418 plus HAT	0.22	123	1.41 (1)‡

* Average from two separate Scatchard analysis with numbers agreeing within 10% of the mean. For LCaN- α_1 , binding was performed on membranes from clone 162. The total number of binding sites (B_{max}) of LCa.11 differs from that of clone 162 at a significance level of $P < 0.001$.

† Average of densities exhibited by cells with net inward currents at +20 to +40 mV (0.5 μ M Bay K8644) of > 5 pA, using 115 mM Ba²⁺ as the charge carrier. For LCaN- α_1 , current density measurements are from 1 to 2 cells each of 13 independent cell clones, including clone 162.

‡ Current density (pA/pF ± s.d. (cells measured)) was recorded using Ba²⁺ as charge carrier at 40 mM instead of 115 mM.

(Table 2). LCaN- α_1 clone 162, which expressed about one third as many specific DHP binding sites as the LCa.11 cell (Table 1), seemed to produce currents most frequently (Fig. 2a) and was used for further transfection studies.

For expression of the rabbit skeletal muscle DHPR- β subunit, its cDNA⁴ was subcloned into the *Eco*RI site of p91023(B) (ref. 12) in which the DHPR- β cDNA is under the control of the adenovirus major late promoter. The plasmid was designated p9-DHPR. β and was transfected with the herpes simplex virus thymidine kinase gene^{9,13} into clone 162 cells. Cells potentially expressing both DHPR α_1 and β messenger RNAs were designated LCaN- $\alpha_1\beta$ cells or $\alpha_1\beta$ cells.

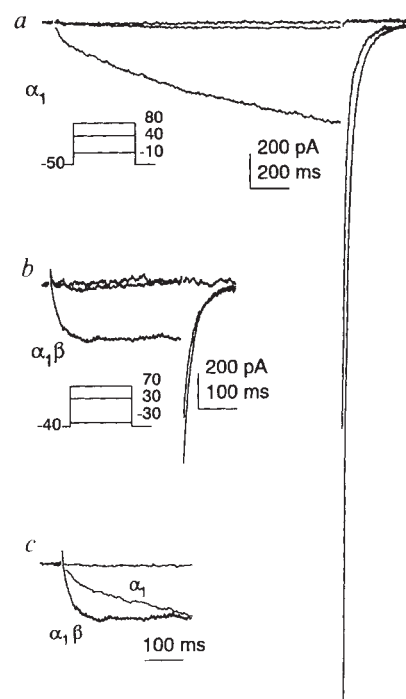
Transformants expressing thymidine kinase and neomycin resistance were selected by growing cells in hypoxanthine/aminopterin/thymidine (HAT) medium with G418 and were cloned using standard techniques. Individual cell clones

were then tested for DHPR- β expression in three assays: northern blot hybridization analysis to detect DHPR β mRNA, specific binding of DHP to their membranes to measure possible effects on the DHP receptor activity of α_1 , and whole-cell patch-clamp measurements¹⁴ to test for possible effects of β on the Ca²⁺ currents produced by α_1 .

Northern blot hybridization of total RNA derived from ten independent neomycin-resistant, thymidine kinase-positive clones revealed that six of the clones were expressing both α_1 and β (Fig. 1a). Coincident with expression of DHPR β mRNA, binding assessed with 0.32 nM [³H](+)-PN200-110, a dihydropyridine Ca²⁺ channel antagonist, was increased by a factor of four to five (Fig. 1b). Scatchard analysis of DHP binding to assess the kinetic basis of this effect revealed that DHPR β increases the maximal number of detectable binding sites without altering their affinity.

FIG. 2 Comparison of Ca²⁺ currents of an α_1 cell (a, LCaN- α_1 clone 162) and from an $\alpha_1\beta$ cell (b, LCaN- $\alpha_1\beta$ clone 7). Note change in timescale between α_1 (a, 200 ms) and $\alpha_1\beta$ (b, 100 ms) records. For both cells the currents were produced by voltage steps to the potentials indicated in the voltage protocol. The potentials correspond to potentials at threshold, maximum inward current and strong depolarization. Currents for α_1 were obtained with the 115 mM Ba²⁺ bath solution and for $\alpha_1\beta$ with 40 mM Ba²⁺ solution. The lower record (c) shows on the same timescale currents from the α_1 cell elicited by potential steps to -10 and 40 mV superimposed on the current from the $\alpha_1\beta$ cell at a step potential of 30 mV. Current amplitudes were normalized to the current at the end of the step from the $\alpha_1\beta$ cell.

METHODS. Cells were grown under standard conditions in 35-mm Falcon Petri dishes on 10-mm square glass coverslips coated with poly-D-lysine in α MEM supplemented with 10% fetal calf serum in the presence of HAT and/or 400 μ g ml⁻¹ G418 as indicated. Seeding was at 5,000 to 10,000 cells per ml and recordings were obtained 3 days afterwards. In the case of LCa.11 cells, 100 μ M ZnSO₄ was added 15 h before use. Coverslips with attached cells to be patched were broken into fragments and the fragments kept at room temperature in Tyrode's solution until used (<1 h). The coverslip fragments were then placed into a static bath containing 0.5 ml of bath solution (composition in mM: BaCl₂ 115, HEPES 10, EGTA 1, pH 7.4, adjusted with *N*-methyl-D-glucamine (NMG)) with 0.5 μ M racemic Bay K 8644. In a few cases a Cl⁻-free bath solution was used (composition in mM: BaOH 40, tetraethylammonium 30, 4-aminopyridine 5, HEPES 10, pH 7.4, adjusted with methanesulphonic acid containing 0.5 μ M racemic Bay K 8644). This did not significantly reduce leak currents and its use was discontinued. Coverslip fragments were discarded after 15 min in bath solution. The pipette filling solution was usually (in mM): HEPES 150, EGTA 20, NMG 110, MgCl₂ 2, pH 7.3, adjusted with NMG. In some experiments a pipette-filling solution of composition (in mM): NMG 130, EGTA 15, BAPTA 5, HEPES 10, MgCl₂ 11.5, CaCl₂ 8, Na₂ATP 3, Na₂GTP 0.1, pH 7.3, adjusted with methanesulphonic acid, was used. There was no consistent difference between currents obtained with these solutions and results have been pooled. Whole-cell patch clamp recordings¹⁴, were performed with a LIST EPC-7 amplifier. Patch clamp electrodes were made from 8161 glass tubing (Garner Glass) and had resistances of 2 M Ω or less. Series resistance error was reduced by adjusting the series resistance compensation circuit of the EPC-7 to just



below the point of oscillation. Whole-cell patch-clamp current data were analysed with pCLAMP software and hardware (Axon Instruments) installed in an IBM-compatible computer. Data were recorded at various digitizing rates and filtered with an eight-pole Bessel filter at 0.2 times the digitizing frequency, except when data were digitized at two frequencies in the same record. Cell capacitance was estimated by applying symmetric linear positive and negative potential ramps under voltage clamp and measuring the capacitive current produced by the ramps.

Stable coexpression of the β component of relative molecular mass 55,000 (M_r 55K) of the rabbit skeletal muscle DHP binding complex resulted in a dramatic increase in the rate of activation of α_1 -mediated currents produced by voltage clamp depolarization (Fig. 2) in three $\alpha_1\beta$ clones selected for electrophysiological studies on the basis of DHP binding (Table 1). Thus, we observed in all three of the selected $\alpha_1\beta$ clones that depolarizations from holding potentials between -50 and -40 mV to test potentials between 10 and 60 mV produced activation rates that were about 100 times faster than those for recipient clone 162, in which α_1 alone expressed Ca^{2+} currents. Despite the large difference in their absolute values, the voltage-dependence of the two sets of rate constants was similar. Deactivation rates measured at the holding potentials were also significantly increased but only by a factor of two (Table 2), far less than the differences observed at depolarized potentials.

We also found slowly activating currents in clone 23 of the LCaN- $\alpha_1\beta$ series. This cell line had undergone the same transfection, selection and cloning procedures as clones 4, 7 and 13, but did not express detectable DHPR β mRNA (Fig. 1). This result makes it unlikely that the accelerated response to depolarization was due to insertional mutagenesis or to non-specific activation of the murine β gene. Using a monoclonal antibody raised against purified rabbit DHPR- β (P.R. and F.H., unpublished results), we found that all clones positive for β mRNA were also positive for expression of an immunoreactive DHPR β protein with the expected apparent M_r of 55K when electrophoresed on polyacrylamide gels (data not shown).

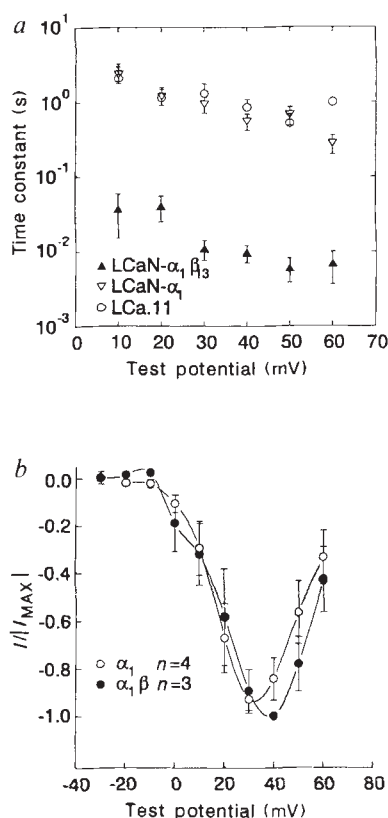


FIG. 3 *a*, Activation time constants for α_1 and $\alpha_1\beta$ currents compared at test potentials between 10 and 60 mV. Activation was fitted to a monoexponential function using nonlinear least-squares optimization. Data were from LCa.11, LCaN- α_1 and LCaN- $\alpha_1\beta$ clone 13. Values at each point are averages from 2-5 cells (LCa.11), 5-11 cells (LCaN- α_1) and 2-4 cells (LCaN- $\alpha_1\beta$). *b*, Comparison of peak current-voltage relationships for α_1 and $\alpha_1\beta$ currents. Peak currents in each cell were normalized by the magnitude of the maximum inward current recorded in the cell. For α_1 currents the potential at which the peak current occurred varied slightly. Both sets of data were measured in 115 mM Ba^{2+} .

TABLE 2 Activation and deactivation times (τ) of murine L cells transfected with the α_1 and β components of the rabbit skeletal muscle DHP receptor complex

Cell line	Deactivation τ (-50 mV) (ms $\pm$ s.d.)	Activation τ (40 mV) (ms $\pm$ s.d.)
LCa.11	59 $\pm$ 14 (n=5)	738 $\pm$ 266 (n=5)
LCaN- α_1 *	63 $\pm$ 9 (n=5)	552 $\pm$ 431 (n=10)
LCaN- $\alpha_1\beta$, clone 13	18 $\pm$ 4 (n=9)	8 $\pm$ 1 (n=5)
LCaN- $\alpha_1\beta$, clone 4	24 $\pm$ 7 (n=5, -40 mV)	5
LCaN- $\alpha_1\beta$ †, clone 7	25 $\pm$ 3 (n=2)	13 $\pm$ 1 (n=2)

* Kinetic measurements are from 1 to 2 cells, each of 9 independent cell clones, including clone 162.

† Kinetic measurements were obtained using Ba^{2+} as charge carrier at 40 mM instead of 115 mM.

thermore, both the recipient clone 162 cell and transfected cell clones that were negative for β mRNA expression were also negative for expression of immunoreactive material.

Although the normalization of activation produced by $\alpha_1\beta$ is clear, the $\alpha_1\beta$ currents were detected less frequently. Gigaseals were more difficult to obtain, and of those that were successful only 28% (32/121) had detectable tail currents. Of this group, only 9/32 showed inward currents of 5 pA or more, even in the presence of Bay K8644 at 0.5 μM . For this reason the results for α_1 and $\alpha_1\beta$ were all obtained using the same concentration of Bay K8644. Under the same conditions, steady-state activation of the coexpressed $\alpha_1\beta$ currents was unchanged from that of the currents expressed by α_1 alone. In two experiments, the potential dependence of steady-state activation was compared between α_1 and $\alpha_1\beta$ using tail current amplitude measurements. The Boltzmann fits to the normalized data had slope factors of 5.2 and 7.0, indicating no change in voltage sensitivity. More compelling was a comparison of the peak current-voltage relationships for four α_1 cells and three $\alpha_1\beta$ cells. The normalized peak current-voltage relationships were similar in the two cases (Fig. 3*b*). As the peak currents correlate with steady-state activation in these non-inactivating currents, we conclude that activation gating is the same for α_1 and $\alpha_1\beta$ cells.

On the basis of the almost 100-fold increase in the rate of activation following its coexpression with α_1 , we propose that the β subunit of the DHPR combines with α_1 as a true subunit of the skeletal muscle Ca^{2+} channel. Such a large effect is unlikely to be due to the presence of Bay K8644. This compound is known to enhance activation and inactivation kinetics of Ca^{2+} currents, possibly by shifting gating to more negative potentials. It also slows deactivation¹⁵ and produces a similar panoply of effects on skeletal muscle Ca^{2+} currents^{16,17}. These effects are much smaller, however, and cannot explain the very large differences in activation rates. Moreover, the β effect on the rate of activation was selective, being much larger at depolarized potentials. As in this case, the deactivation rate for Ca^{2+} currents measured at potentials below threshold is faster than the activation rates measured at depolarized potentials (Table 2). The deactivation rate is mainly voltage-independent¹⁸, which is consistent with the idea that the final closed-open transition in voltage-sensitive Na^+ , K^+ and Ca^{2+} channels is voltage-independent. Thus one possibility is that the β subunit has an unusually pronounced effect on voltage-dependent transitions among earlier closed states of the channel; perhaps this is related to the lower current densities produced in the $\alpha_1\beta$ cells.

At present we cannot explain the discrepancy between the increased DHP binding capacity of $\alpha_1\beta$ cells and the lack of significant change in current densities. A similar discrepancy has been noted before for DHPR in skeletal muscle¹⁹. It may be that inactivation of channels formed by $\alpha_1\beta$ is increased, although 30-s hyperpolarizing prepulses to -100 mV did not

affect test current amplitudes. These prepulses might not have been sufficient to correct an ultraslow inactivation process. Another factor could be a reduction in mean open time. It may also be that increased DHP binding occurs in cells where $\alpha_1\beta$ does not reach functional maturity as an ion channel, and hence that changes in DHP binding at the cell population level do not reflect the binding properties that exist in the individual cells from which Ca^{2+} channel currents are recorded.

Our experiments not only demonstrate the functional interaction between α_1 and β , but they also raise questions concerning the roles of γ and $\alpha_2\delta$ in Ca^{2+} channel function, and the extent to which these are tissue-specific. Several α_1 genes with varying tissue-specific expression, each giving rise to more than one alternatively spliced product, have been identified²⁰. Likewise, northern analysis indicates the existence of more than one form of the β subunit^{4,10}. No γ -like mRNA has been found in non-skeletal muscle tissues^{7,8} and $\alpha_2\delta$ is expressed in all tissues that express α_1 (ref. 3). The γ subunit may be a skeletal muscle-specific component responsible for the high levels of expression of DHP binding sites in this tissue. Also, Mikami *et al.*²¹ have found that co-injection of mRNA encoding cardiac α_1 with mRNA encoding skeletal muscle $\alpha_2\delta$ into *Xenopus* oocytes results in a roughly twofold increase in current densities over those obtained with α_1 mRNA alone. It will be of interest to see whether stable expression of all skeletal muscle DHPR

components in L cells produces current densities much larger than have so far been obtained with α_1 or with α_1 plus β . □

Received 18 March; accepted 25 June 1991.

1. Catterall, W. A. *Science* **242**, 50–61 (1988).
2. Tanabe, T. *et al. Nature* **328**, 313–318 (1987).
3. Ellis, S. B. *et al. Science* **241**, 1661–1664 (1988).
4. Ruth, P. *et al. Science* **245**, 1115–1118 (1989).
5. Jay, S. D. *et al. Science* **248**, 490–492 (1990).
6. Bosse, E. *et al. FEBS Lett.* **267**, 153–156 (1990).
7. DeJongh, K. S., Warner, C. & Catterall, W. A. *J. biol. Chem.* **265**, 14738–14741 (1990).
8. Jay, S. D. *et al. J. biol. Chem.* **266**, 3287–3293 (1991).
9. Perez-Reyes, E. *et al. Nature* **340**, 233–236 (1989).
10. Kim, H. S. *et al. J. biol. Chem.* **265**, 11858–11863 (1990).
11. Nakada, T., Mishina, M. & Numa, S. *FEBS Lett.* **211**, 5–9 (1987).
12. Wong, G. G. *et al. Science* **228**, 810–815 (1985).
13. Liao, C.-F. *et al. J. biol. Chem.* **264**, 7328–7337 (1989).
14. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
15. Lacerda, A. E. & Brown, A. M. *J. gen. Physiol.* **93**, 1243–1273 (1989).
16. Cognard, C., Traore, F., Potreau, D. & Raymond, G. *Pflügers Arch. ges. Physiol.* **407**, 677–638 (1986).
17. Cognard, C., Romey, G., Galizzi, J.-P., Fosset, M. & Lazdunski, M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1518–1522 (1986).
18. Brown, A. M., Tsuda, Y. & Wilson, D. L. *J. Physiol.* **344**, 549–583 (1983).
19. Schwartz, L. M., McCleskey, E. W. & Almers, W. *Nature* **314**, 747–750 (1985).
20. Perez-Reyes, E., Wei, X., Castellano, A. & Birnbaumer, L. *J. biol. Chem.* **265**, 20430–20436 (1990).
21. Mikami, A. *et al. Nature* **340**, 230–233 (1989).

ACKNOWLEDGEMENTS. Supported in part by the National Institutes of Health (A.M.B., E.P.-R. and L.B.) and the Deutsche Forschungsgemeinschaft (F.H.). Parts of this research will serve towards the fulfilment of requirements for a Ph.D. degree (H.S.K.).

Kinetic maturation of an immune response

Jefferson Foote & César Milstein

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Is the affinity maturation of antibodies under thermodynamic or kinetic control, or both? We compared the physical constants of hapten binding by antibodies from 2-phenyl-5-oxazolone-specific hybridomas from primary, secondary and tertiary responses. In addition to an increase in equilibrium constant, there was a shift in the antibody repertoire after the primary response towards an immunoglobulin family with an extremely high on-rate constant. This shift occurred in spite of the average or below-average affinity of this group of antibodies. This is consistent with B-lymphocyte proliferation being subject to a kinetic selection, with a premium on binding target antigens rapidly, in parallel with a thermodynamic selection based on binding tightly.

The early immune response to 2-phenyl-5-oxazolone (phOx)¹ is dominated by a single germline V gene in the κ locus, designated V_{κ} -Ox1, and a heavy chain V gene, V_H -Ox1 (refs 2–4). After prolonged immunization, two maturation strategies are apparent⁵. The V_{κ} -Ox1 and V_H -Ox1 genes appear as derivatives with point mutations, ranging from two per chain at early stages³ to eight in the one-year tertiary response⁶, representing a process of mutational 'drift'. A repertoire 'shift' is manifested in the rise of phOx-specific antibodies, of which the light chain, heavy chain, or both, are derived from germline V genes non-cognate with V_{κ} -Ox1 or V_H -Ox1. Both cognate and noncognate antibodies generally have higher antigen affinity than the archetype V_{κ} -Ox1/ V_H -Ox1. Two subsets become apparent; one consists of derivatives of V_{κ} -Ox1 in association with V_H -Ox1, but also with heavy chains belonging to other gene families. Of the mixed Ox1/non-Ox1 category, the pattern of conservation of key residues suggests that in spite of the use of a different V gene, the folding of the loops of the complementarity-determining regions recreates a facsimile of the V_{κ} -Ox1/ V_H -Ox1 antigen binding site⁷. The second subset expresses light chains that are apparent derivatives of a gene designated $V_{\kappa}45.1$. The

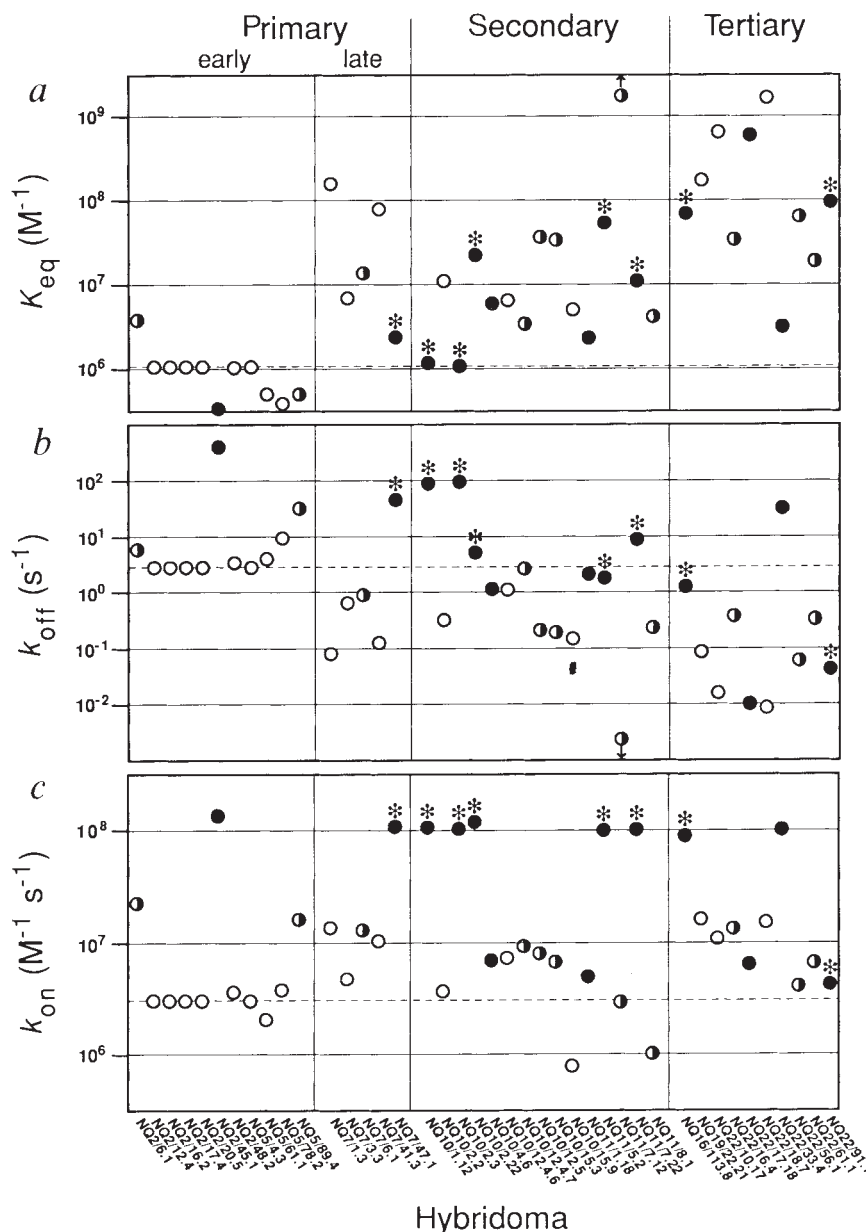
corresponding heavy chains are also quite different from the previous subset, and although probably not derived from a single germ-line gene, are closely related members of the TEPC15 (group 7 according to Dildrop⁸) gene family. The $V_{\kappa}45$ /TEPC15 combination was absent at seven days, but in the secondary response the numbers of this subset in the population were near parity with Ox1 (ref. 5). The two subsets represent distinct solutions to the architectural problem of binding phOx.

Concurrent equilibrium and kinetic measurements were made on purified antibodies, using the γ -amino butyrate conjugate of phOx (phOx-GABA) and identical temperature and solution conditions (Fig. 1). The previously noted progressive increase in affinity accruing during the maturation of the immune response is evident from the logarithmic representation of equilibrium constants in Fig. 1a. Antibodies from fusions seven days after immunization had weak affinities for phOx-GABA, which at about 10^6 M^{-1} is a value marginally lower than that previously derived using phOx-amino-caproic acid². Antibodies from later-stage fusions had an affinity increased by 1–3 orders of magnitude. These antibodies were not only derived by point mutation of the Ox1 archetype: non-Ox1-cognate molecules also showed improved affinities relative to the early primary response. But within any stage, antibodies arising by the shift strategy showed no superiority in affinity over those modified through drift. In the secondary response, the equilibrium constants of cognate and noncognate molecules were in the same numerical range; in the tertiary response, values for the noncognate examples were mostly lower. Thus affinity-driven selection, although clearly operative, was insufficient to account for the pattern of gene usage in later stages of the anti-phOx response. We sought another selective factor that might underlie the rise of the non-cognate antibodies, in particular the $V_{\kappa}45$ /TEPC15 subset.

When the process of maturation was studied at the kinetic level, the off-rate (Fig. 1b) resembled the equilibrium data in Fig. 1a, inasmuch as the diagonally downward trend in off-rate, corresponding to a longer lifetime of the hapten-antibody complex, matched the upward trend in affinity. This accords with ideas on the kinetic parameters underlying affinity maturation. Non-Ox1-cognate antibodies tend to display faster off-rates, but the existence of a systematic quantitative kinetic difference distinguishing the Ox1 and $V_{\kappa}45$ /TEPC15 subsets was most

FIG. 1 Physical constants of the pHx-GABA-antibody reaction. Hybridoma lines are listed at the bottom, and the constant measured with the corresponding antibody plotted directly above. Vertical lines delimit the data according to the stage of the immune response from which the hybridomas were derived. Early and late primary refer to cell fusions at 7 and 14 days after immunization; secondary refers to fusions after a booster injection at 6 weeks, and tertiary to fusions at 1 year after a second booster^{2,3,5,6}. ○, Antibodies of which heavy and light chains are both derived from the OX1 germ-line genes; ●, both chains derived from other V gene families; ◐, mixed category. *a*, Equilibrium (association) constants. *b*, First-order rate constants for dissociation; these were not measured directly, but calculated from the equilibrium constant and forward rate constant, *c*, Second-order rate constants for association. Data points representing antibodies composed of a V κ -45.1 light chain and TEPC15 heavy chain are marked with an asterisk. Of the remaining non-OX1-cognate hybridomas, NQ2/45.1, NQ10/4.6 and NQ22/33.4 also use a V κ -45.1 light chain. The atypical kinetic behaviour of NQ22/91.1 may be attributed to recombination of the TEPC15 V gene with an unusual D-segment. For further sequence information, see refs 2, 3, 5 and 6. The affinity of NQ11/7.12 was beyond the sensitivity of our method of measurement; arrows in *a* and *b* indicate that the value is off-scale. Five independent hybridoma lines obtained in the early primary response (NQ2/20.5, NQ2/12.4, NQ2/16.2, NQ2/17.4, NQ5/4.3) produce covalently identical, canonical OX1 antibodies. Measurements were made only on NQ2/20.5, but data for the others were presumed to be identical, and are listed here to give a more accurate reflection of the day-7 repertoire. A dashed line is drawn through these points in each part of the figure to facilitate comparison with this benchmark. The standard errors of the equilibrium constant determinations were ~5% for most points; only two high-affinity antibodies exceeded 20%. In kinetic measurements, standard errors of the group of antibodies having k_{on} near $10^8 \text{ M}^{-1} \text{ s}^{-1}$ clustered around 8%, with NQ2/45.1 as an outlier; the remainder gave standard errors that were on average 5%.

METHODS. The hapten pHx-GABA was synthesized by the method used for the analogous amino-caproyl compound¹, and twice recrystallized in the acid form from dimethylformamide/water. Under the conditions of the measurements, pHx-GABA is a 76:18:6 mixture of structures. The two main species are related by a moderately rapid ($0.1\text{--}1 \text{ s}^{-1}$) conformational isomerization at the enamine nitrogen, and the third is related to the other two by configurational isomerization at the methynyl carbon¹⁵. Thus, although the empirical constants we report, which are derived from least-squares fits of experimental data to simple kinetic and equilibrium models, accurately reflect the partition of pHx-GABA between the bound and free states, they are not strictly true physical constants characteristic of a discrete chemical reaction. All hybridoma lines secreted immunoglobulins with κ light chains and IgG1 heavy chains, except for NQ11/5.2, which was of the IgG2a isotype. Antibodies were purified to homogeneity from hybridoma culture medium or ascites fluid by affinity chromatography on Protein A-Sepharose. Immunoglobulin concentrations were determined by spectrophotometry. A valency of two was assumed for those molecules whose weak affinities precluded direct measurement of valency by hapten titration (see below). All measurements were taken at 20 °C in 25 mM NaH₂PO₄, 125 mM NaCl, 0.2 mM EDTA buffer, pH 7. Affinity and second-order rate constants were determined by fluorescence techniques, based on the quenching of tryptophan fluorescence by the bound hapten¹⁶. To optimize the wavelength selection, the emission spectra of 100 nM antibody samples with and without 1 μM pHx-GABA were scanned in a Perkin-Elmer LS-5B spectrofluorimeter using an excitation wavelength of 280 nm. A difference spectrum was calculated in each case, and the wavelength of maximum Δ fluorescence was noted for subsequent measurements. A regime of hapten excess was used for K_{eq} determination of relatively weak-affinity antibodies. Negligible volumes of hapten were added to a 3-ml antibody sample in a cuvette in the instrument to cover a concentration range of 1/4 to 4 times the preliminary estimate of the dissociation constant, and the fluorescence determined after each addition. The antibody concentration was chosen to be at most 1/5, on a site basis, of the lowest hapten concentration. The value of the equilibrium constant was obtained from a least-squares fit of the data to a hyperbola. For a tight-binding regime, antibody at a concentration on a site basis roughly 10 times the preliminary estimate of the dissociation constant was titrated to twofold excess with hapten and sample fluorescence was determined 2 min after each addition. Fluorescence data were fitted by least-squares to the equation: $F =$



$F_0 r A - (f/2)((rA + L + K) - ((rA + L + K)^2 - 4rAL)^{1/2})$, where F is the observed fluorescence, A is the antibody concentration, L the hapten concentration, F_0 the molar (per site) antibody fluorescence, r the number of binding sites per antibody, f the molar fluorescence change and K the equilibrium constant, expressed in this case (but not in the figure) as a dissociation constant. The derivation of a similar binding equation is given in ref. 17. The four latter parameters are optimized in the fitting routine, using the Marquadt algorithm¹⁸, and reported with standard errors. If the antibody concentration used was below 100 nM, 30 $\mu\text{g ml}^{-1}$ ubiquitin was added as a carrier. As ubiquitin lacks tryptophan¹⁹, it is very weakly fluorescent, and its use minimizes artefactual quenching. Rate constants for association were determined using a stopped-flow instrument⁹ (Hi-Tech Scientific) operated in fluorescence format with 280 nm excitation, and the emission observed using a filter opaque below 320 nm. To avoid the complicated direct analysis of a second-order reaction, rate determinations were made of a series of pseudo-first order reactions, this simplification being made feasible by use of hapten concentrations in ≥ 10 -fold molar excess over binding sites²⁰. Portions (50 μl) of 100–250 nM antibody were mixed with equal volumes of hapten, and the resulting reaction, which seldom deviated from a simple exponential, was followed for about six half-lives. Multiple identical runs were averaged, and the averaged signal fit by least-squares to obtain a pseudo-first-order rate constant. These apparent values obtained at four or five different hapten concentrations were plotted against the final hapten concentration after mixing, and the slope of a regression line was taken as the second-order rate constant for association.

TABLE 1 Activation energy for binding pHx-GABA

Hybridoma	V _K	V _H	E_{act} (kcal mol ⁻¹)	k_{on} (M ⁻¹ s ⁻¹)
NQ2/12.4	Ox1	Ox1	7.0	3.0×10^6
NQ10/15.9	Ox1	Ox1	9.5	7.9×10^5
NQ10/12.5	Ox1	MOPC21	9.0	8.0×10^6
NQ16/113.8	V _K 45.1	TEPC15	4.6	9.0×10^7
NQ22/18.7	Ox1	Ox1	9.8	1.5×10^7

NQ10/12.5 is an antibody of known three-dimensional structure. NQ10/15.9 and NQ16/113.8 were chosen to compare extremes of k_{on} . NQ2/12.4 is the canonical Ox1 form with no point mutations, whereas NQ22/18.7 is a high-affinity, highly mutated derivative. The values of k_{on} at 20 °C (k_{20°) are shown for reference. E_{act} , activation energy. Kinetic measurements are described in the legend to Fig. 1 and were taken at 5° intervals from 278K to 313K. The second-order rate constants obtained were used in an Arrhenius plot ($\ln k_{on}$ against $1/T$) and the activation energy taken from the slope of a regression line through these points.

apparent from comparison of on-rate constants (k_{on}) (Fig. 1c). Rather than an overt diagonal pattern, the points are segregated into two groups, separated along the ordinate by an order of magnitude. V_K45/TEPC15 antibodies had an association rate constant of 10^8 M⁻¹ s⁻¹, an essentially invariant value, despite mutations that greatly altered affinity. Thus the major new component of the anti-phOx repertoire emerging in the secondary and tertiary responses, although offering no particular advantage in affinity, shows an extremely fast on-rate constant.

The association rate constants of antibodies in the Ox1 group covered a broad range of values. The Ox1 prototype, NQ2/20.5, showed a low on-rate of 3×10^6 M⁻¹ s⁻¹. Higher values generally arose during maturation—very few points fall below the dashed line in Fig. 1c marking the canonical antibody. But the increase in k_{on} did not go beyond just over 10^7 M⁻¹ s⁻¹, a tenth of that of the V_K45/TEPC15 group. The Ox1 antibodies observed in the tertiary response, which are heavily mutated and have very high affinity, fail to exceed this limit, nor is it breached by antibodies with either the heavy or light chain replaced with a non-Ox1-cognate partner. Thus the drift maturation strategy did not apparently yield an antibody with the distinctive fast kinetics of the V_K45/TEPC15 group. A structural constraint appears innate to the geometry of the Ox1-type binding site, limiting the maximum association rate to a value far lower than that which, given the V_K45/TEPC15 examples, is physically possible for combination of a molecule the size of pHx-GABA with an immunoglobulin.

A set of anti-phOx antibodies was chosen for further study. The temperature dependence of reaction with pHx-GABA revealed a large and systematic difference in Arrhenius activation energy (Table 1), distinguishing the very fast V_K45/TEPC15 antibody from those in the Ox1 group. The former, NQ16/113.8, showed an activation energy of 4.6 kcal mol⁻¹, which is characteristic of a diffusion-controlled reaction in an aqueous solvent and may be attributed entirely to the effect of changes in the viscosity of water with temperature⁹. The other entries all show considerably higher activation energies; thus binding of hapten to an Ox1-type site entailed overcoming an energy barrier which was absent in the V_K45/TEPC15 example.

The early belief that hapten binding by antibodies is uniformly fast^{9,10} was premature¹¹. Indeed, the range of on-rate constants of anti-phOx antibodies, even in the handful of hybridomas obtained from a single mouse, is virtually as broad as the spread of all reported values^{12,13}. Furthermore, systematic shifts towards higher k_{on} associated with a repertoire shift indicate a biological significance for this parameter; whether the selective recruitment of memory cells or the selection within germinal centres of newly mutated lymphocytes is affected, remains obscure.

Kinetic analysis of the population dynamics of the anti-phOx

repertoire leads us to propose that the parameter k_{on} is a critical factor in antigen-driven B-cell selection. In addition to the phenomenon of affinity maturation, a parallel process of kinetic maturation occurs, defined as the antigen-driven mutational refinement of the antibody repertoire to give net increases in bimolecular antibody-antigen association rate constant. This observation is of obvious relevance to attempts to imitate (or even improve) *in vivo* maturation using bacterial expression systems¹⁴. The appearance of point mutations in the Ox1 archetype often coincides with a small increase in k_{on} . But there is a structural constraint on k_{on} , detectable as an activation energy barrier (Table 1), which may be an intrinsic property of the deep and constricted binding pocket of the Ox1 type of antibody⁷. Inability to escape this constraint through point mutation thus potentiates the shift in the later anti-phOx repertoire, inexplicable on affinity grounds, to the kinetically superior V_K45/TEPC15 geometry. □

Received 1 March; accepted 17 June 1991.

1. Mäkelä, O., Kaartinen, M., Pelkonen, J. L. T. & Karjalainen, K. *J. exp. Med.* **148**, 1644–1660 (1978).
2. Kaartinen, M., Griffiths, G. M., Markham, A. F. & Milstein, C. *Nature* **304**, 320–324 (1983).
3. Griffiths, G. M., Berek, C., Kaartinen, M. & Milstein, C. *Nature* **312**, 271–275 (1984).
4. Even, J., Griffiths, G. M., Berek, C. & Milstein, C. *EMBO J.* **4**, 3439–3445 (1985).
5. Berek, C., Griffiths, G. M. & Milstein, C. *Nature* **316**, 412–418 (1985).
6. Berek, C., Jarvis, J. M. & Milstein, C. *Eur. J. Immunol.* **17**, 1121–1129 (1987).
7. Alzari, P. M. *et al.* *EMBO J.* **9**, 3807–3814 (1990).
8. Dildrop, R. *Immun. Today* **5**, 85–86 (1984).
9. Day, L. A., Sturtevant, J. M. & Singer, S. J. *Ann. N.Y. Acad. Sci.* **103**, 611–625 (1963).
10. Sehon, A. H. *Ann. N.Y. Acad. Sci.* **103**, 626–630 (1963).
11. Denizot, F. C., Hirn, M. H. & DeLaage, M. A. *Molec. Immunol.* **16**, 509–513 (1979).
12. Smith, T. W. & Skubitz, K. M. *Biochemistry* **14**, 1496–1502 (1975).
13. Mason, D. W. & Williams, A. F. in *Handbook of Experimental Immunology*, Vol. 1, 4 edn (eds Weir, D. M., Herzenberg, L. A., Blackwell, C. & Herzenberg, L. A.) 38.1–38.17 (Blackwell Scientific, Oxford, 1986).
14. Winter, G. & Milstein, C. *Nature* **349**, 293–299 (1991).
15. Titman, J. J., Foote, J., Jarvis, J., Keeler, J. & Neuhaus, D. *Chem. Commun.* 419–421 (1991).
16. Eisen, H. N. *Meth. med. Res.* **10**, 115–121 (1964).
17. Segal, I. *Enzyme Kinetics* 73–74 (Wiley-Interscience, New York, 1975).
18. Marquardt, D. W. *J. Soc. Ind. Appl. Math.* **11**, 431–441 (1963).
19. Schlesinger, D. H. & Goldstein, G. *Nature* **255**, 423–424 (1976).
20. Benson, S. W. *The Foundations of Chemical Kinetics* 81–82. (McGraw-Hill, New York, 1960).

ACKNOWLEDGEMENTS. We thank J. M. Jarvis for purifying many of the antibodies used in this work. J.F. was supported by the Jane Coffin Childs Memorial Fund for Medical Research and a Merck Fellowship.

Breakdown of self-tolerance in anergic B lymphocytes

Christopher C. Goodnow*, Robert Brink & Elizabeth Adams

Centenary Institute for Cancer Medicine and Cell Biology, University of Sydney, New South Wales 2006, Australia

* Howard Hughes Medical Institute and Department of Microbiology and Immunology, Beckman Center, Stanford University, California 94305, USA

PRODUCTION of autoantibodies, which characterizes most autoimmune diseases, is normally avoided by active elimination^{1–7} or functional inactivation (anergy)^{8–15} of B and T lymphocytes bearing receptors for self antigens. The mechanisms leading to the escape of self-reactive clones from these normal tolerance mechanisms in autoimmune diseases nevertheless remain obscure. Here, we demonstrate that clonal anergy in B lymphocytes is a reversible process, and that silenced self-reactive B cells can be reactivated under particular conditions to give rise to vigorous antibody responses. Reactivation of anergic lymphocytes may explain many examples of transient autoimmune reactions in normal individuals, and may under pathological conditions be important in the development of chronic autoimmune disease.

Previous studies of 'double-transgenic' mice expressing rearranged anti-lysozyme immunoglobulin genes and a gene construct encoding lysozyme have documented the intrinsic

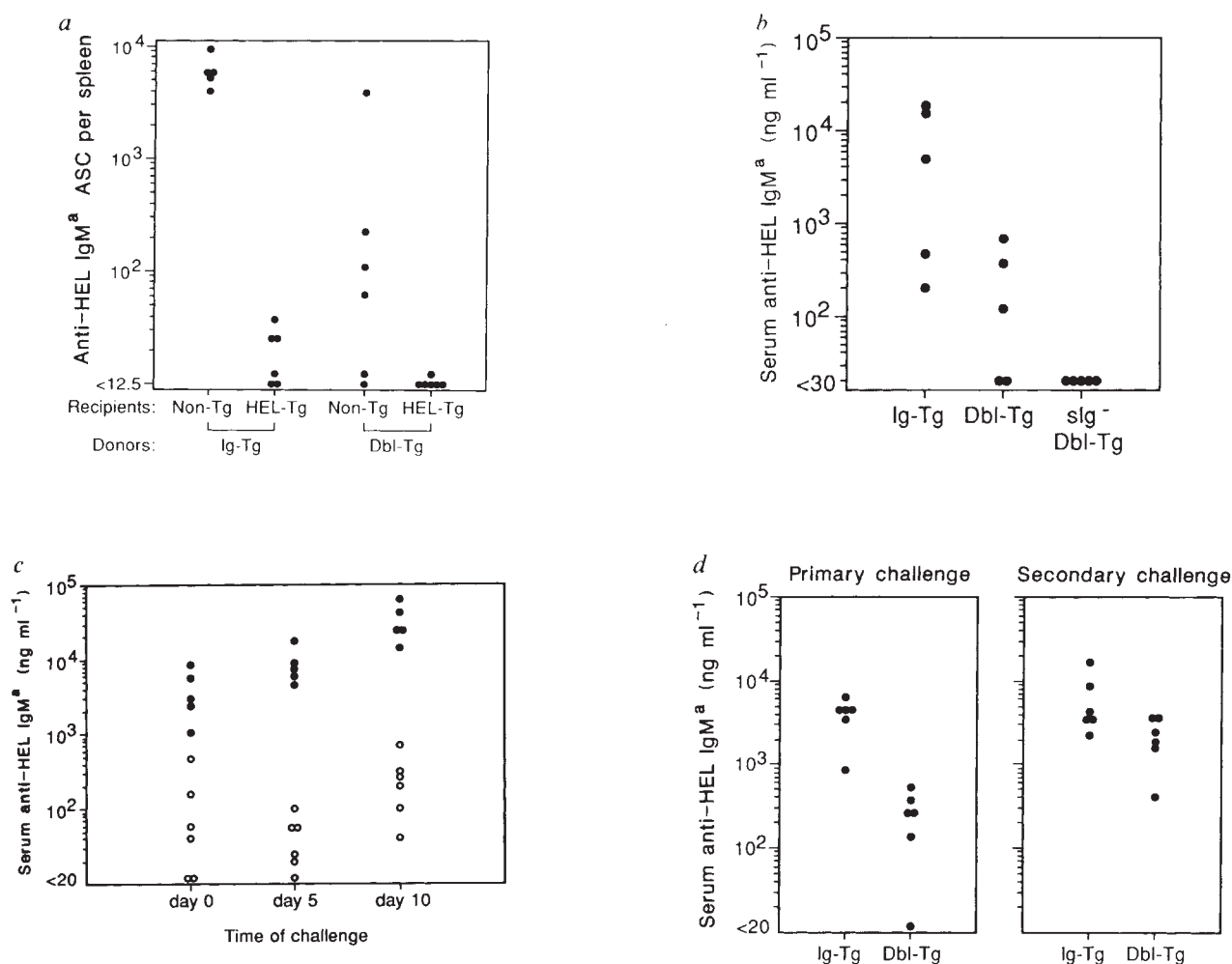


FIG. 1 Recovery of function in tolerant B cells *in vivo*. *a*, Antibody response from immunoglobulin-transgenic (Ig-Tg) or double-transgenic (Dbl-Tg) B cells following transfer into irradiated nontransgenic (Non-Tg) or lysozyme-transgenic (HEL-Tg) recipients. Spleen cells from nontransgenic mice primed previously to sheep red blood cells (SRBC) were cotransferred as a source of helper T cells, and the recipients challenged at the time of transfer with lysozyme coupled to SRBC. Antibody responses were measured 5 days later by enumerating cells secreting lysozyme-binding IgM of the α -(transgene-encoded) allotype in the spleen of individual recipients (indicated by dots). *b*, Absence of adoptive antibody response from double-transgenic spleen cells depleted of surface-immunoglobulin bearing cells (slg⁻ Dbl-Tg) before transfer and challenge in nontransgenic recipients. *c*, Antibody response from immunoglobulin (●) or double-transgenic (○) B cells after transfer into irradiated nontransgenic recipients as in *a*, except that the recipients were challenged either at the time of transfer, or after waiting 5 or 10 days. *d*, Restimulation of adoptive transfer recipients. Nontransgenic recipients of immunoglobulin- or double-transgenic spleen cells were challenged as in *a* with lysozyme conjugated to horse red blood cells at the time of transfer and the antibody response measured in the serum 7 days later (primary response). After the serum

antibody titres had returned to background, the animals were immunized a second time with lysozyme conjugated to sheep red blood cells and the antibody response was again measured in the serum 7 days later (secondary response).

METHODS. Spleen cells from male C57BL/6-derived MD3 × ML5 immunoglobulin- and double-transgenic mice, or nontransgenic C57BL/6 mice primed to SRBC or horse red blood cells (HRBC), were prepared as described previously⁹. Transgenic spleen cells (10^6 , *a*, *c* or 2×10^4 , *b* or 5×10^3 , *d*) were injected intravenously (i.v.), together with 5×10^6 nontransgenic primed cells, into 750 rad-irradiated male C57BL/6-derived nontransgenic or ML5 HEL-transgenic recipients as indicated. *b*, slg-bearing cells were depleted with anti-immunoglobulin-coated magnetic beads as described previously¹⁷. Recipients were challenged by i.v. injection of 2×10^8 HEL-SRBC or HEL-HRBC. *d*, Restimulation was performed 5 months after the initial transfer with 2×10^8 HEL-SRBC given intraperitoneally. Responses were measured 5 days after challenge (*a*) by enumerating anti-lysozyme IgM⁺-secreting cells in the spleen of individual recipient animals by Spot-ELISA¹⁶, and 7 days after challenge (*b*, *c* and *d*) by measuring anti-lysozyme IgM^a in serum by ELISA¹⁶.

ance of tolerance in the former group and partial recovery of B-cell function in the absence of continuous exposure to lysozyme in the latter recipients. Importantly, antibody production by double-transgenic cells in nontransgenic recipients was completely abolished if surface immunoglobulin-bearing B cells were first depleted from the transferred inoculum (Fig. 1*b*), thereby excluding the possibility that the antibody response was due to differentiation of nontolerant pre-B cells after transfer.

Although tolerant B cells removed from the double-transgenic mice clearly regained considerable ability to mount an antibody response after a single round of *in vivo* stimulation, the fact that the response remained lower than that of nontolerant controls suggested that a certain period of time was required for complete functional recovery. To test this possibility, double-transgenic B cells and nontolerant immunoglobulin-transgenic

functional silencing of lysozyme-reactive B cells as a result of encountering a critical level of autologous lysozyme *in vivo*^{9,16,17}. To explore the alternative possibilities that anergic B cells were reversibly or irreversibly silenced, B cells from double-transgenic mice and nontolerant immunoglobulin-transgenic controls were transferred, together with helper T cells specific for foreign red blood cells, into either nontransgenic mice or into lysozyme-transgenic littermates expressing tolerogenic concentrations of serum lysozyme. Both groups of recipients were then challenged with lysozyme conjugated to foreign red blood cells to elicit a collaborative immune response between the transferred B and T cells. No anti-lysozyme antibody was produced by the double-transgenic B cells in lysozyme-transgenic recipients, whereas a significant response did occur when the cells were transferred into nontransgenic animals (Fig. 1*a*), consistent with mainten-

controls were 'parked' in a resting state in irradiated nontransgenic hosts for 0, 5 or 10 days before challenging the recipients. We have previously found that expression of membrane IgM but not IgD was reduced 10–20-fold in the double-transgenic B cells⁹ and that this downregulation of IgM was closely correlated with tolerance induction¹⁶. Consistent with gradual recovery, fluorescence-activated cell sorter (FACS) analysis of the parked double-transgenic B cells revealed spontaneous recovery of membrane IgM expression on the tolerant cells, reaching amounts identical to that on parked B cells from immunoglobulin-transgenic controls after 10 days (Fig. 2a). The increased expression of membrane IgM was due to recovery

from downregulation rather than to selection of rare IgM^{high} variant cells or to *de novo* maturation of initially surface immunoglobulin-negative pre-B cells, as lysozyme-binding cells exhibited intermediate levels of IgM expression at earlier time-points (Fig. 2a) and were undetectable in recipients of immunoglobulin- or double-transgenic spleen cells which had first been depleted of surface immunoglobulin-bearing cells before transfer (not shown). The immunoglobulin- and double-transgenic B cells also had an equivalent ability to home to the spleen and persist for at least 10 days in the recipient animals (Fig. 2b).

Unexpectedly, the parked double-transgenic B cells

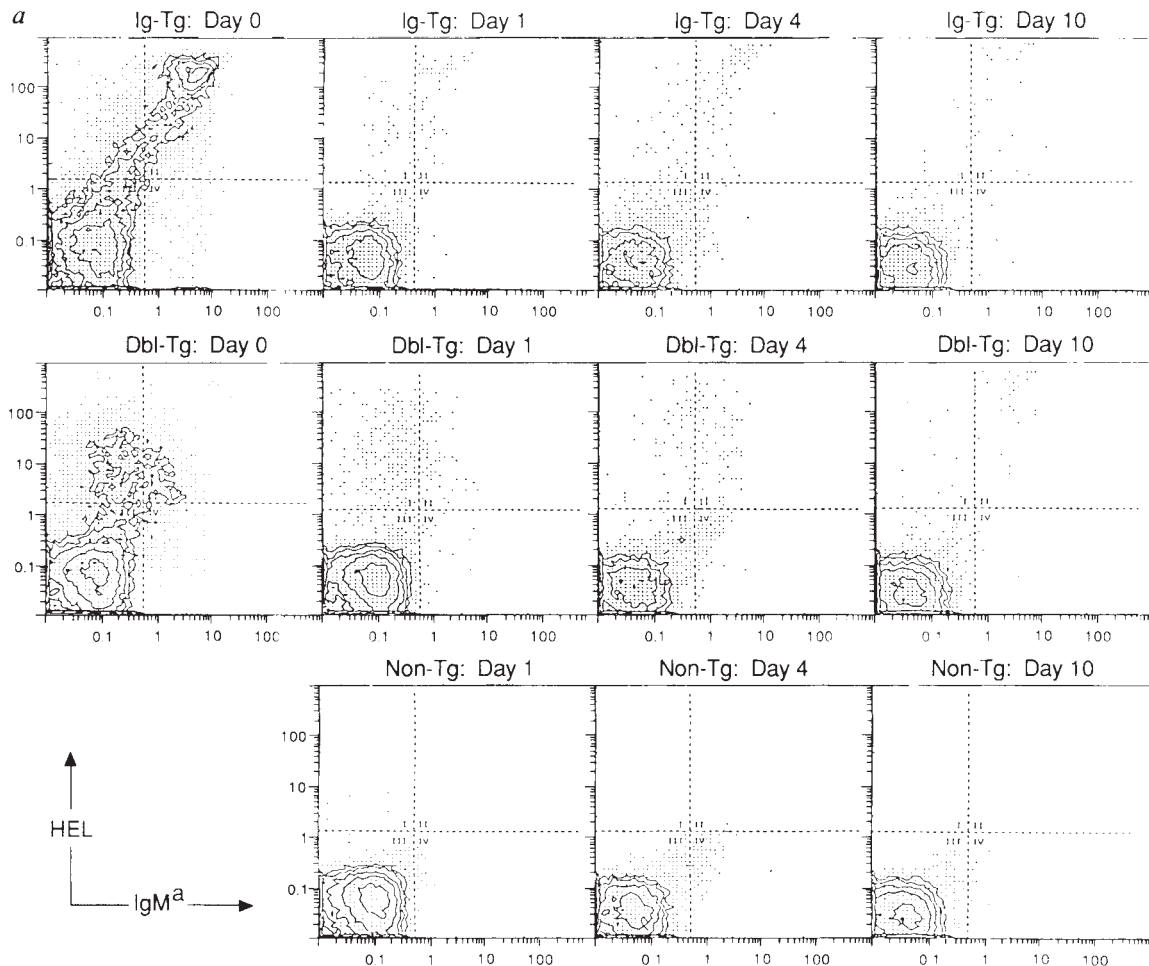
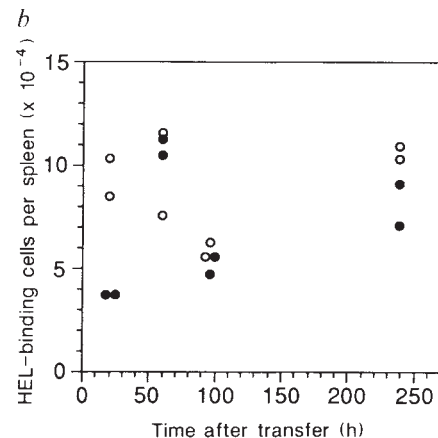


FIG. 2 Recovery of membrane IgM expression and persistence of tolerant B cells 'parked' in nontransgenic recipients. Immunoglobulin-transgenic or double-transgenic spleen cells (denoted day 0) were injected intravenously into irradiated nontransgenic recipients. At 1, 4 and 10 days after injection, lysozyme-binding B cells expressing transgene-encoded IgM (IgM^a) were detected in the spleen of individual recipients by immunofluorescence and FACS analysis. *a*, IgM^a-expression on lysozyme-binding cells from double-transgenic mice (dots represent 1 or more cells per 4 × 4 channel grid), compared with transferred immunoglobulin-transgenic or nontransgenic controls. *b*, Numbers of lysozyme-binding cells present in the spleen of individual recipients of Ig-transgenic (●) or double-transgenic (○) donor cells at different times after transfer.

METHODS. MD3 × ML5 immunoglobulin- or double-transgenic spleen cells (10⁷), or nontransgenic C57BL/6 controls, were injected i.v. into 750 rad-irradiated C57BL/6 nontransgenic recipients. Spleen cell suspensions were prepared from individual recipients at the times indicated, stained for lysozyme binding and IgM^a-expression, and analysed by FACS, as described previously⁹. The number of lysozyme-binding cells per spleen was calculated by multiplying the percentage of lysozyme binding cells by the total number of spleen cells. The number of lysozyme-binding B cells transferred to each recipient mouse initially was 185 × 10⁴ (Ig-Tg) or 104 × 10⁴ (Dbl-Tg).



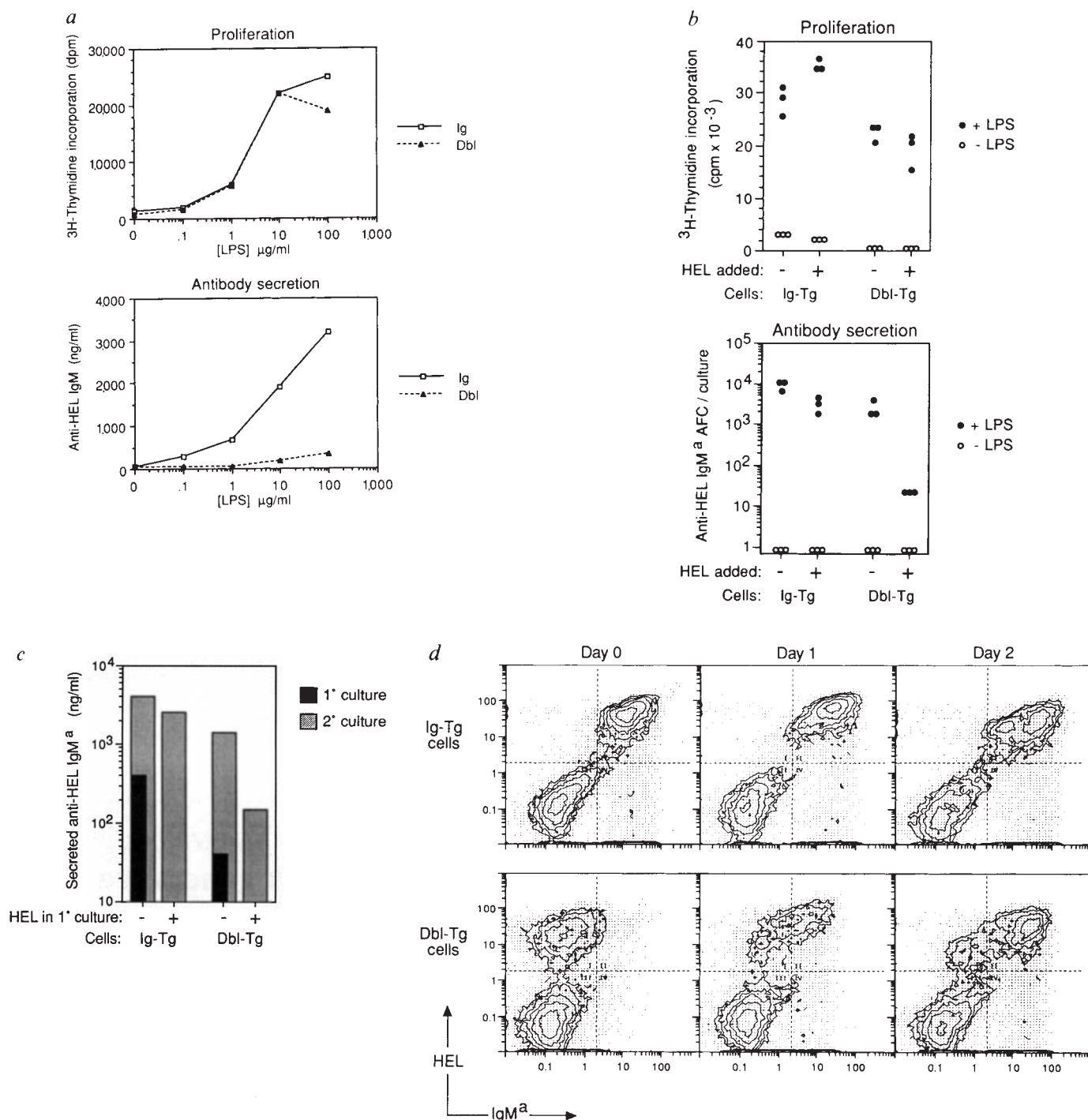


FIG. 3 Recovery of tolerant B cells *in vitro*. **a**, Immunoglobulin- or double-transgenic spleen cells were stimulated into mitosis with LPS, and proliferation measured on the third day by thymidine incorporation. Secreted anti-lysozyme antibody in the culture supernatants was quantitated at the same time by ELISA. **b**, Equivalent stimulation of immunoglobulin- or double-transgenic B cells, but performed in the presence or absence of 100 ng ml^{-1} lysozyme added to the culture medium. Proliferation was measured after 1 day, and antibody secretion measured after 3 days by enumerating cells secreting anti-lysozyme IgM. **c**, Gradual recovery of antibody secretion in mitogenically-stimulated double-transgenic B cells. Immunoglobulin or double-transgenic spleen cells were stimulated for 3 days in the presence or absence of lysozyme as in **b**, then washed and replated at the starting cell density in fresh cultures without lysozyme. Antibody secretion during the first 3-day culture (primary culture) and after reculture for a further 3 days (secondary culture) was quantitated by ELISA. **d**, Recovery of membrane IgM expression on stimulated double-transgenic B cells. Immunoglobulin- or double-transgenic spleen cells were stimulated with LPS for the times

indicated and then analysed for lysozyme-binding and α -allotype membrane IgM by FACS.

METHODS. Spleen cells from immunoglobulin- or double-transgenic mice, and from C57BL/6 and BALB/c nontransgenic controls (not shown) were cultured at 10^5 cells per well in 96-well flat-bottomed plates as described previously¹⁷. Lysozyme was added to a final concentration of 100 ng ml^{-1} at the start of the cultures, and LPS was added at $20 \mu\text{g ml}^{-1}$ unless otherwise shown. For FACS analysis, the cells were collected from the wells after 20 or 43 h by Pasteur pipette, washed twice in 2% fetal calf serum in phosphate buffered saline, counted and stained for lysozyme binding and IgM^a expression as described previously⁸. Parallel staining with anti-B220 demonstrated that blastogenesis was confined to B cells in the cultures. Thymidine incorporation was measured as described⁸ by pulsing the cultures for 12 h with $0.5 \mu\text{Ci } ^3\text{H}$ -thymidine starting either at 17 or 60 h of culture. Antibody-secreting cells were enumerated by ELISA-spot assays done on individual cultures after washing the collected cells three times to remove any remaining lysozyme.

nevertheless showed no evidence for spontaneous functional recovery (measured by antibody production) during the period before stimulation. Thus when the transfer recipients were challenged with lysozyme conjugated to foreign red blood cells to initiate collaboration between red cell-specific helper T cells and the lysozyme-specific B cells, an equivalent low-level antibody response was mounted by the 0, 5 or 10-day parked double-transgenic cells (Fig. 1c). The same results were obtained with a variety of antigen doses and routes of administration. Over and above the spontaneous recovery of membrane IgM expression observed in the unstimulated parked B cells, it thus appeared that recovery of effector function required the reversal of additional intracellular events and was only initiated once the B cells were stimulated through interaction with helper T cells. Consistent with this possibility, much greater antibody secretion could be elicited from the double-transgenic B cells by administering a second booster immunization to the transfer recipients (Fig. 1d), the resulting serum concentrations being only twofold lower than those in boosted recipients of non-tolerant immunoglobulin-transgenic controls.

A very similar pattern of functional recovery was observed in double-transgenic B cells stimulated to divide *in vitro* by the bacterial mitogen lipopolysaccharide (LPS). Tolerant B cells removed from the double-transgenic environment and cultured with LPS in the absence of lysozyme were stimulated to divide as efficiently as nontolerant immunoglobulin-transgenic B cells, but secreted much less antibody initially (Fig. 3a). After several days of LPS-stimulated mitogenesis in culture, however, the number of anti-lysozyme antibody-secreting cells and amount of secreted antibody gradually approached that in parallel cultures of immunoglobulin-transgenic cells (Fig. 3b,c). Membrane IgM expression also recovered progressively on the entire population of LPS-stimulated double-transgenic B cells, attaining levels equivalent to those on LPS-stimulated immunoglobulin-transgenic B cells within 2 days (Fig. 3d). No functional or phenotypic recovery was observed in parallel cultures lacking LPS.

In a manner also comparable to that observed in the transfer experiments *in vivo*, constant exposure to lysozyme *in vitro*, by including lysozyme in the culture medium at concentrations equal to those present *in vivo* completely prevented functional recovery of the tolerant double-transgenic B cells. Although proliferation of the double-transgenic B cells was unaffected by the presence of lysozyme (Fig. 3b), differentiation of the B cells into antibody-secreting cells was abolished (Fig. 3b). Moreover, tolerant B cells driven through multiple rounds of cell division over 3 days in the presence of lysozyme secreted antibody as inefficiently as freshly collected B cells after washing and reculture (Fig. 3c), whereas identical cells stimulated for 3 days in the absence of lysozyme secreted antibody after reculture at levels approaching those of the nontolerant immunoglobulin-transgenic controls.

Taken together, the reactivation of tolerant B cells observed *in vivo* and *in vitro* in this study demonstrates that functionally silenced self-reactive B lymphocytes are not irreversibly committed to unresponsiveness and eventual cell death, and that these cells do not appear to have an intrinsically shortened lifespan. The dissociation between recovery of antigen-receptor expression and recovery of the ability to mount an antibody response (Figs 1 and 2), points to other changes in the tolerant B cells in addition to the previously established connection between induction of tolerance and receptor downregulation¹⁶. The fact that recovery of antibody secretion appeared to require multiple rounds of mitogenic stimulation, either by interaction with helper T cells or by stimulation with LPS, and was abolished if the cells continued to bind lysozyme, may suggest the existence of a repressor system which is induced and maintained by antigen-receptor signals and is inactivated or diluted out during mitosis. The reversible nature of tolerance in anergic B lymphocytes has potentially important consequences both for normal

B cell physiology and in autoimmunity. During the course of a physiological immune response, it is conceivable that silenced self-reactive B cells which crossreact with foreign antigens may be activated by helper T cells and adopt a more appropriate specificity by undergoing immunoglobulin-gene hypermutation¹⁸. In autoimmunity, on the other hand, initially silenced self-reactive B cell clones may be pathologically triggered to produce antibody by a shift in the balance between tolerance, reinforced by continued binding of self antigens, and activation by potent costimuli, such as bacterial endotoxins¹⁹⁻²¹ or helper T cells specific for self or foreign antigens²²⁻²⁴. □

Received 18 February; accepted 20 June 1991.

1. Burnet, F. M. *The Clonal Selection Theory of Acquired Immunity* (Vanderbilt University Press, Nashville, 1959).
2. Kappler, J. W., Roehm, N. & Marrack, P. *Cell* **49**, 273-280 (1987).
3. Kappler, J. W., Staerz, U., White, J. & Marrack, P. C. *Nature* **332**, 35-40 (1988).
4. MacDonald, H. R. *et al. Nature* **332**, 40-45 (1988).
5. Kisielow, P., Blüthmann, H., Staerz, U. D., Steinmetz, M. & von Boehmer, H. *Nature* **333**, 742-756 (1988).
6. Sha, W. C. *et al. Nature* **336**, 73-76 (1988).
7. Nemazee, D. A. & Bürki, K. *Nature* **337**, 562-566 (1989).
8. Nossal, G. J. V. *A. Rev. Immun.* **1**, 33-62 (1983).
9. Goodnow, C. C. *et al. Nature* **334**, 676-682 (1988).
10. Rammensee, H.-G., Kroschewski, R. & Frangoulis, B. *Nature* **339**, 541-544 (1989).
11. Ramsdell, F., Lantz, T. & Fowlkes, B. J. *Science* **246**, 1038-1041 (1989).
12. Burky, L. C., Lo, D., Kanagawa, O., Brinster, R. L. & Flavell, R. A. *Nature* **342**, 564-566 (1989).
13. Bonneville, M. *et al. Nature* **344**, 163-165 (1990).
14. Blackman, M. A. *et al. Nature* **345**, 540-542 (1990).
15. Schwartz, R. H. *Science* **248**, 1349-1356 (1990).
16. Goodnow, C. C., Crosbie, J., Jorgensen, H., Brink, R. A. & Basten, A. *Nature* **342**, 385-391 (1989).
17. Adams, E., Basten, A. & Goodnow, C. C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5687-5691 (1990).
18. Naparstek, Y. *et al. J. exp. Med.* **164**, 614-626 (1986).
19. Primi, D., Hammarstrom, L., Smith, C. I. E. & Moller, G. *J. exp. Med.* **145**, 21-30 (1977).
20. Izui, S., Lambert, P. H., Fournie, G., Turler, H. & Miescher, P. A. *J. exp. Med.* **145**, 1115 (1977).
21. Hang, L. *et al. J. exp. Med.* **157**, 874-883 (1983).
22. Gleichmann, E. & Gleichmann, H. *Eur. J. Immun.* **6**, 899-906 (1976).
23. Gleichmann, E., Van Elven, E. H. & Van der Veen, J. P. W. *Eur. J. Immun.* **12**, 152-159 (1982).
24. Weigle, W. O. *Adv. Immun.* **30**, 159-273 (1980).

ACKNOWLEDGEMENTS. We thank A. Basten for generous support and critical comments on the manuscript and H. Jorgensen for performing FACS analyses. The research was funded by the NHMRC and the Australian Research Council.

The mdx mouse diaphragm reproduces the degenerative changes of Duchenne muscular dystrophy

H. H. Stedman, H. L. Sweeney, J. B. Shrager, H. C. Maguire, R. A. Panettieri, B. Petrof, M. Narusawa, J. M. Leferovich, J. T. Sladky & A. M. Kelly

Schools of Medicine and Veterinary Medicine, University of Pennsylvania, 3800 Spruce Street, Philadelphia, Pennsylvania 19104, USA

ALTHOUGH murine X-linked muscular dystrophy (mdx) and Duchenne muscular dystrophy (DMD) are genetically homologous and both characterized by a complete absence of dystrophin^{1,2}, the limb muscles of adult mdx mice suffer neither the detectable weakness nor the progressive degeneration that are features of DMD³⁻⁸. Here we show that the mdx mouse diaphragm exhibits a pattern of degeneration, fibrosis and severe functional deficit comparable to that of DMD limb muscle, although adult mice show no overt respiratory impairment. Progressive functional changes include reductions in strength (to 13.5% of control by two years of age), elasticity, twitch speed and fibre length. The collagen density rises to at least seven times that of control diaphragm and ten times that of mdx hind-limb muscle. By 1.5 years of age, similar but less severe histological changes emerge in the accessory muscles of respiration. On the basis of these findings, we propose that dystrophin deficiency alters the threshold

for work-induced injury. Our data provide a quantitative framework for studying the pathogenesis of dystrophy and extend the application of the mdx mouse as an animal model.

Histopathological changes in the mdx diaphragm are rare before 25 days of age. At 30 days, foci of myofibre degeneration, necrosis, mineralization and regeneration are present but are not extensive. Similar changes are present in limb muscles. We find no evidence that the initial lesions of the diaphragm either precede or are more serious than those in other muscles of the mdx mouse.

In contrast to mdx limb muscles, in which regeneration largely restores muscle structure (Fig. 1a), the diaphragm undergoes progressive degeneration. At 6 months of age there is wide variation in myofibre size and architecture, continued necrosis and significant connective tissue proliferation (Fig. 1c). Evidence of regenerative activity persists as most fibres have central nuclei at this age. By 16 months, the entire mdx diaphragm is grossly pale owing to extensive myofibre loss and replacement fibrosis (compare Fig. 1b, control, and 1d, mdx). As a result, all of the remaining fibres are ensheathed in dense collagenous tissue and most have architectural aberrations. The collagen density, as estimated from hydroxyproline measurements⁹, rises to at least seven times that of control diaphragm and ten times that of mdx hind-limb muscle (Fig. 1f).

Measurable changes in the physiological properties of the

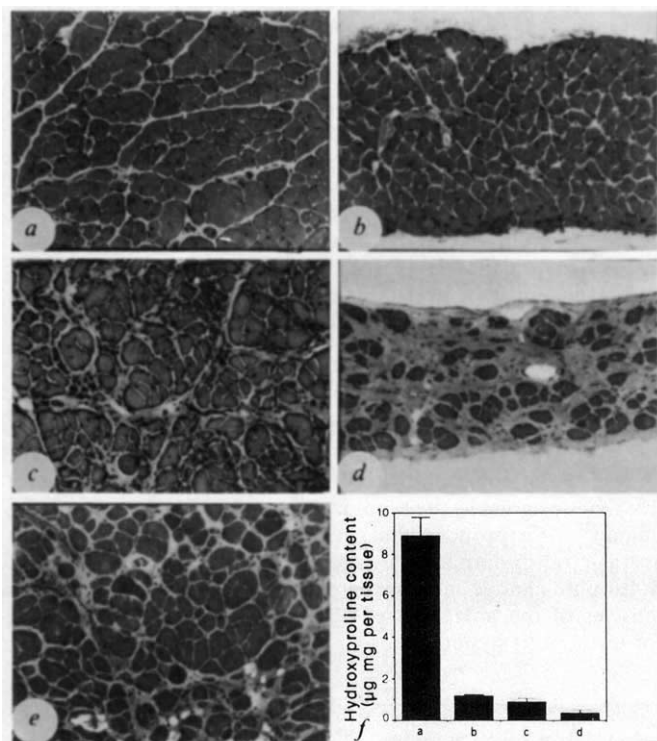


FIG. 1 Haematoxylin and eosin-stained transverse sections through frozen sections of skeletal muscles of mdx and control C57Bl10 mice. Magnification, 100 $\times$. a, Tibialis anterior at 16 months, mdx, note central myonuclei in most fibres; b, diaphragm at 16 months, control; c, diaphragm at 6 months, mdx; d, diaphragm at 16 months, mdx; e, intercostal muscle at 16 months, mdx; f, hydroxyproline contents of mouse muscles, normalized to unit tissue wet weight. a, Mdx diaphragm; b, control diaphragm; c, mdx quadriceps; d, control quadriceps. Column heights and error bars represent means and standard deviations, respectively; $n=4-6$ each.

METHODS. Hydroxyproline contents of diaphragm muscle strips and quadriceps muscle fragments dissected free of tendinous tissue were determined by a colorimetric method⁹. Tissue fragments were blotted dry and weighed immediately before lyophilization. Median age of mice was 16 months, range 12-18 months. Controls were age- and sex-matched C57Bl10 mice.

mdx diaphragm parallel those detectable histologically, and represent the first progressive changes documented in this animal model. By 16 months, the proportion of mdx diaphragm fibres staining immunochemically for slow myosin¹⁰ has doubled relative to that of the age-matched control. This resembles the changes reported in DMD¹¹. In the same mdx muscles the kinetics of isometric force generation have slowed relative to controls by a factor of 4 (Fig. 2). Tetanic stimulation of diaphragm muscle strips from 16-22-month-old control mice yields a maximal isometric strength of 27.4 ± 3.5 g mm⁻² (mean $\pm$ s.d., $n=10$ each). By contrast, isometric strength in the mdx mouse diaphragm drops to 5.7 ± 1.2 and 3.7 ± 0.72 g mm⁻² at 16 and 22 months, respectively. In these dystrophic muscles, the mean fibre length at L_0 has shortened to $65\% \pm 2\%$ of the control value.

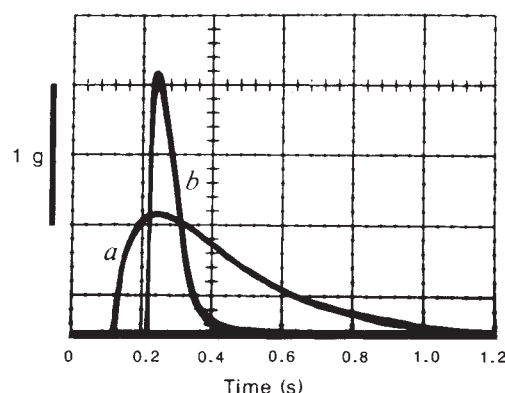
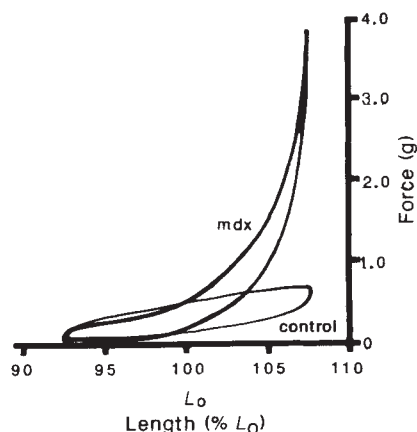


FIG. 2 Isometric twitch kinetics for mdx (a) and control (b) mouse diaphragm muscle strips. Time intervals elapsing between the onset of tension and the achievement of maximal tension are 120 ms for a and 30 ms for b. Intervals for decay of tension from maximal to half-maximal are 280 ms for a and 55 ms for b, respectively. Cross-sectional areas, calculated as described below, were 1.2 mm² and 0.54 mm² for the mdx and control muscle strips, respectively. Normalized to these cross-sectional areas, the ratio of peak twitch forces for this pair is $\sim 1:5$.

METHODS. Freshly resected diaphragms were submerged in oxygenated Ringer's solution buffered to pH 7.5 with 25 mM HEPES. One muscle strip ~ 5 mm wide, centered on the insertion of the phrenic nerve, was isolated from each hemidiaphragm. Each strip was mounted horizontally by pinning the superficial periosteum of several ribs to a mobile platform and the central tendon to the lever of a Cambridge Technology (Watertown MA) model 300 H motor servosystem. Stimulation was induced by passing current between platinum wire electrodes mounted 1 mm above and below the mid-portion of each strip. A series of twitch forces generated at incrementally different fibre lengths was used to identify the point of optimal sarcomere length for tetanic force measurement (L_0). Twitches were recorded on a Techtronix model 5441 storage oscilloscope at a sweep rate of 200 ms per division. The figure represents a pair of storage oscilloscope traces, superimposed against a common scaling grid. Instrument gain settings were identical for both twitches. Values reported here and in the text represent averages of five measurements made directly from the oscilloscope screen. The average length of the muscle fibres at L_0 was estimated by viewing each strip in this configuration against a 0.1 mm scale. Strips were then tetanized at 65 Hz, the minimal frequency that induces twitch fusion at room temperature. All measurements were made at $23 \pm 1^\circ$. Following dissection free of tendinous attachments, the muscular component of each strip was weighed to the nearest 0.1 mg. Force values were normalized to calculated mean cross-sectional areas at L_0 , assuming a tissue density of 1 g cm⁻³. To compare sarcomere lengths, diaphragmatic strips were held at L_0 and cooled to 0 $^\circ$ C by recirculating chilled Ringer's solution. Fixation was obtained in this configuration by the addition, with rapid mixing, of glutaraldehyde/paraformaldehyde from a concentrated stock to 4% w/v. The temperature was maintained at 0 $^\circ$ C for at least 10 min. Fixed samples were embedded in plastic and sectioned longitudinally. Sarcomere lengths were determined for at least 10 separate fibres from each of four fixed specimens by light microscopic examination against a micrometre scale. Mdx and sex-matched C57Bl10/ScSn control were both 1.5 years old.



Direct measurements of sarcomere length at this muscle fibre length demonstrate values of $\sim 2.5 \mu\text{m}$ for both mdx and control diaphragm strips, indicating that, by this stage, the average mdx diaphragm myofibril has undergone a 35% net loss in sarcomeres.

As in DMD, compositional changes in the mdx diaphragm are accompanied by a severe loss of tissue compliance (Fig. 3). In a diaphragm strip from a 1.5-year-old mdx mouse, passive stretch over the range of $L_0 \pm 7\%$ induces a rapid increase in stiffness beginning at $\sim 97\%$ of L_0 . The dynamic elastic modulus of this tissue is immeasurably low at 93% of L_0 , increases to 15 g mm^{-2} at L_0 , and to more than ten times that at $L_0 + 7\%$. A control strip from an age-matched nondystrophic mouse has an elastic modulus of less than 5 g mm^{-2} over this range of lengths. Only at a length of $L_0 + 15\%$ did the elastic modulus approach that of the dystrophic muscle at L_0 (data not shown).

The striking differences between mdx diaphragm and limb muscle could reflect intrinsic differences in regenerative capacity. But immunochemical staining of fibres with a monoclonal antibody against embryonic myosin and the presence of basophilic fibres with central nuclei indicate that attempted regeneration persists in the mdx diaphragm even at 16 months. Electron microscopic examination reveals that $\sim 6\%$ of peripheral muscle nuclei in the normal mature mouse diaphragm reside in satellite cells ($n=500$). This compares with 5% of peripheral nuclei in the mature soleus, a slow muscle of the limb, and 2% in the fast-twitch extensor digitorum longus muscle^{12,13}. *In vitro* studies of satellite cells from these three sources reveal similar proliferative capacities¹⁴.

The ageing mdx mouse shows no overt respiratory compromise, despite progressive diaphragm weakness. Dystrophic changes appearing in the anterior scalene and intercostal muscles between 1 and 1.5 years of age resemble those of the diaphragm at six months of age (Fig. 1c and e). These findings suggest that compensatory recruitment of the accessory respiratory muscle mass eventually occurs, and that the transferred ventilatory workload accelerates the degenerative process in these muscles as well. The measurable increase in stiffness associated with progressive diaphragm fibrosis should preclude overstretch injury in the muscle fibres remaining at this age. In this sense, connective tissue infiltration of the degenerating diaphragm represents an adaptive response. Moreover, additional collagen support may be unusually important for the diaphragm as sarcomere length changes during respiration are generally greater than those in limb muscle during locomotion^{15,16}.

These observations have particular relevance to DMD, in

FIG. 3 Passive hysteresis curves²⁵ for mdx and control mouse diaphragm muscle strips undergoing cyclical length changes centered on L_0 . Both loops are clockwise. Peak-to-peak cycle strains, expressed as percentages of L_0 , are 14% for both muscle strips. At $L_0 + 7\%$, the dynamic elastic modulus of the mdx strip is more than 30 times that of the control strip, as calculated by averaging the slopes of the lengthening and shortening arms of the loops and dividing by the average tissue cross-sectional area. The shapes of the curves were almost invariant over a cycle frequency range of $2\text{--}40 \text{ Hz}$, encompassing the physiological range of respiratory rates in the mouse. Cross-sectional areas for both strips at L_0 are $0.5 \pm 0.02 \text{ mm}^2$.

METHODS. Strips were mounted and force was transduced as described for Fig. 2. Muscle lengths were set at L_0 and cross-sectional areas determined as described⁹. Sinusoidal length perturbations around L_0 were controlled by a Wavetek model 88 signal generator coupled to the servomotor system. The oscilloscope trace was generated by controlling the horizontal axis with the signal generator output and the vertical axis with the servomotor input voltage. The figure represents a pair of storage oscilloscope traces, superimposed against a common grid. Vertical axis instrument gain was the same for both strip recordings. Horizontal gain was adjusted to normalize peak-to-peak trace excursion for muscle strips of different lengths at L_0 . The figure depicts results obtained with strips from 1.5-year-old mice.

which respiratory failure is the leading cause of death. Respiratory muscle degeneration and fibrosis probably occur early in the course of DMD as decrements in maximal inspiratory pressure and vital capacity have been documented by five years of age in asymptomatic boys¹⁷. The present findings emphasize the importance of targeting myoblast¹⁸ or gene^{19,20} replacement therapies towards the respiratory muscles in DMD and provide an appropriate model system for evaluating the functional impact of dystrophin restoration.

Dystrophin-deficient muscle fibres have increased sarcolemmal fragility²¹. Work-induced injury seems to be a normal property of skeletal muscle and may be necessary to induce many of the adaptive responses to exercise^{22,23}. The large reserve of satellite cells in the diaphragm and soleus compared with fast-twitch limb muscles is consistent with this interpretation^{12,13}. Our results raise the possibility that dystrophin deficiency alters the threshold for work-induced injury, thereby perturbing the normal balance between injury and repair in muscle fibres bearing the greatest workload. In contrast to obligatory respiratory work, the power output of limb muscles is minimal in the cage-reared mouse. Moreover, the respiratory work rate per unit mass varies as the negative 0.22 power of the mass in mammals. Hence, the minimal obligatory rate in the 30 g mouse is about five times as great as the basal respiratory work rate in the 70-kg human²⁴. We propose that this scale-related discrepancy in average power output partly accounts for the distinct pattern of dystrophic change seen in the diaphragm compared with limb muscles of the mdx mouse. Our results provide a framework for testing this hypothesis. □

Received 18 March; accepted 28 June 1991.

- Hoffman, E. P., Brown, R. H. & Kunkel, L. M. *Cell* **51**, 919–928 (1987).
- Sicinski, P. et al. *Science* **244**, 1578–1580 (1989).
- Dangain, J. & Vrbova, G. *Muscle & Nerve* **7**, 700–704 (1984).
- Coulton, G. R., Curtin, N. A., Morgan, J. E. & Partridge, T. A. *Neuropath. appl. Neurobiol.* **14**, 53–70 (1988).
- Coulton, G. R., Curtin, N. A., Morgan, J. E. & Partridge, T. A. *Neuropath. appl. Neurobiol.* **14**, 299–314 (1988).
- Anderson, J. E., Bressler, B. H. & Ovalle, W. K. *J. Muscle Res. Cell Motil.* **9**, 499–515 (1988).
- Mandel, J. L. *Nature* **339**, 584–586 (1989).
- DiMario, J., Buffinger, N., Yamada, S. & Strohman, R. C. *Science* **244**, 688–690 (1989).
- Switzer, B. R. & Sumner, G. K. *Analyt. Biochem.* **39**, 487–491 (1971).
- Narusawa, M. et al. *J. Cell Biol.* **104**, 447–459 (1987).
- Webster, C., Silberstein, L., Hays, A. P. & Blau, H. M. *Cell* **52**, 503–513 (1988).
- Kelly, A. M. *Dev. Biol.* **65**, 1–10 (1978).
- Schmalbruch, H. & Hellhammer, U. *Anat. Rec.* **189**, 169–176 (1977).
- Dusterhoft, S., Yablonska-Reuveni, Z. & Pette, D. *Differentiation* **45**, 185–191 (1990).
- Griffiths, R. I. *J. Neurosci. Meth.* **21**, 159–165 (1987).
- Newman, S. et al. *J. appl. Physiol.* **56**, 753–764 (1984).
- Smith, P. E. M. et al. *New Engl. J. Med.* **316**, 1197–1204 (1987).
- Partridge, T. A., Morgan, J. E., Coulton, G. R., Hoffman, E. P. & Kunkel, L. M. *Nature* **337**, 176–9 (1989).

19. Wolff, J. A. *et al. Science* **247**, 1465–1468 (1990).
20. Lee, C. C., Pearlman, J. A., Chamberlain, J. S. & Caskey, C. T. *Nature* **349**, 334–336 (1991).
21. Menke, A. & Jockusch, H. *Nature* **349**, 69–71 (1991).
22. McCully, K. K. & Chance B. in *Respiratory Muscles and Their Neuromotor Control* 317–326 (Liss; New York, 1987).
23. Kasper, C. E., White, T. P. & Maxwell, L. C. *J. appl. Physiol.* **68**, 533–539 (1990).
24. Stahl, W. R. *J. appl. Physiol.* **22**, 453–460 (1967).
25. Syme, D. A. *J. Physiol. Lond.* **424**, 301–315 (1990).

ACKNOWLEDGEMENTS. We thank K. Knight and C. Clark for help with the hydroxyproline assays, K. McKully for help with the muscle strip studies, and E. Jullian and J. Hayden for photographic assistance. This work is supported by grants from the Muscular Dystrophy Association, the Lucille P. Markey Charitable Trust, and from the National Heart, Lung and Blood Institute. This work is dedicated to the memory of Holt and Roland Stedman.

GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families

Andrea Vortkamp, Manfred Gessler
& Karl-Heinz Grzeschik

Institut für Humangenetik, Philipps-Universität, Bahnhofstrasse 7a,
D-3550 Marburg, Germany

THE Greig cephalopolysyndactyly syndrome (GCPS) is an autosomal dominant disorder affecting limb and craniofacial development in humans^{1,2}. GCPS-affected individuals are characterized by postaxial polysyndactyly of hands, preaxial polysyndactyly of feet, macrocephaly, a broad base of the nose with mild hypertelorism and a prominent forehead. The genetic locus has been pinpointed to chromosome 7p13 by three balanced translocations associated with GCPS in different families^{3,4,19}. This assignment is corrobor-

ated by the detection of two sporadic GCPS cases carrying overlapping deletions in 7p13 (ref. 7), as well as by tight linkage of GCPS to the epidermal growth factor receptor gene in 7p12–13 (ref. 8). Of the genes that map to this region, those encoding T cell receptor- γ , interferon- β 2, epidermal growth factor receptor, and *Hox1.4*, a potential candidate gene for GCPS, have been excluded from the region in which the deletions overlap^{7,9}. Here we show that two of the three translocations interrupt the *GLI3* gene, a zinc-finger gene of the *GLI-Krüppel* family already localized to 7p13 (refs 5, 6). The breakpoints are within the first third of the coding sequence. In the third translocation, chromosome 7 is broken at about 10 kilobases downstream of the 3' end of *GLI3*. Our results indicate that mutations disturbing normal *GLI3* expression may have a causative role in GCPS.

For a reverse genetics approach we established a human-mouse somatic cell hybrid panel subdividing the short arm of chromosome 7 into 11 intervals¹⁹. In this panel the GCPS region is defined by two hybrid clones derived from each GCPS translocation cell line retaining one or both of the translocation chromosomes and segregating the intact chromosome 7. Four random probes mapped into the shortest interval encompassing the translocation breakpoints with two probes located on either side (Fig. 1a, b). The translocation breakpoints fall within a 630-kilobase (kb) *NotI* restriction fragment identified by three of the flanking markers¹⁹.

In a parallel approach, we tested a candidate gene sequence by hybridizing a genomic fragment amplified by polymerase chain reaction (PCR) from the central part of *GLI3*, a gene recently mapped to 7p13 by *in situ* hybridization⁶. This gene had been isolated by virtue of its cross-hybridization to the zinc-finger gene *GLI*, which is amplified in certain glioblastomas and represents a potential sequence-specific DNA-binding transcription factor from the *GLI-Krüppel* gene family⁵. The *GLI3*

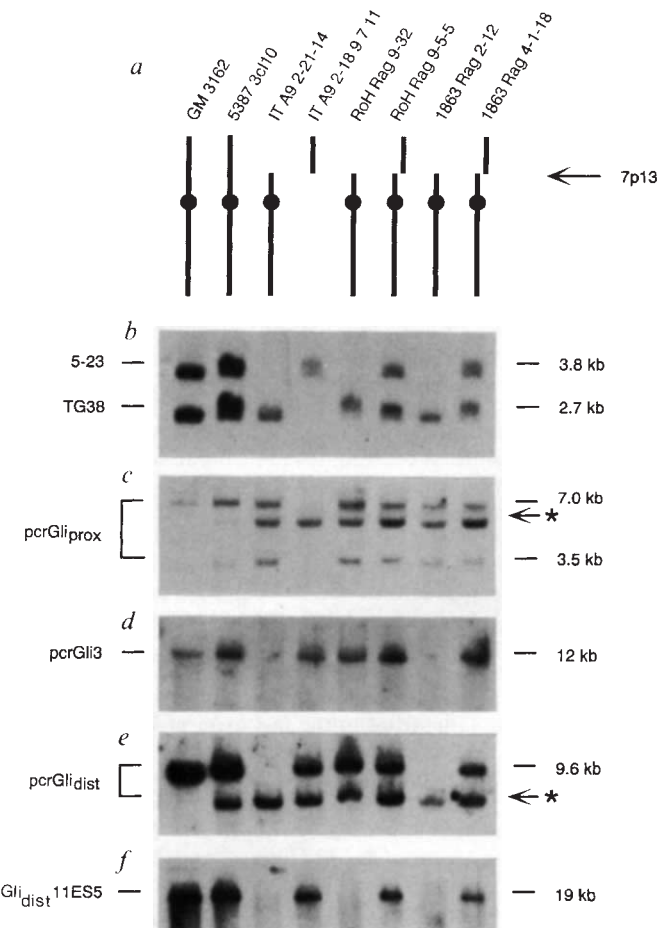


FIG. 1 GCPS translocation breakpoints disrupt the *GLI3* gene in 7p13. *a*, Schematic diagram depicting the presence of human chromosome-7 material retained in different lines of the human-mouse somatic cell hybrid panel¹⁹ (GM3162, human control DNA; 5387-3c10, chromosome 7-only hybrid; IT, RoH and 1863 cell lines, GCPS translocation hybrids). *b–f*, Hybridization of the somatic cell hybrid panel with different probes from 7p13. Control probes TG38 (ref. 15) and 5-23 (ref. 16) map within the closest intervals on the centromeric and telomeric side of the GCPS translocation breakpoints (*b*). Probes from the *GLI3* locus are ordered from centromere to telomere with pcrGli_{prox} (*c*) located centromeric to all three translocations, followed by pcrGli3 (*d*) and pcrGli_{dist} (*e*) mapping between the IT/1863 and RoH breakpoints. Probe Gli_{dist}11ES5 (*f*) is localized telomeric to all translocation breakpoints.

METHODS. Somatic cell hybrid panels (5 μ g control DNA and 15 μ g hybrid DNA, both digested with *EcoRI*) were hybridized and washed under high stringency¹⁷. Cross-hybridizing mouse fragments are marked by asterisks. Probe pcrGli3 was derived by PCR from genomic DNA with primers corresponding to nucleotides 8–27 and 592–610 of pGLI3HH-HaeIII-640 (ref. 5). Probes pcrGli_{prox} and pcrGli_{dist} (nucleotides 98–570 and 4,354–4,933 of the cDNA sequence⁶) were amplified after reverse transcription of adult lung total RNA (First Strand kit, Stratagene) using 20-mer primers. PCR conditions were as follows: 40 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 60 s. Probe Gli_{dist}11ES5 is a 5-kb *SaI* fragment of phage Gli_{dist}11 described in Fig. 2.

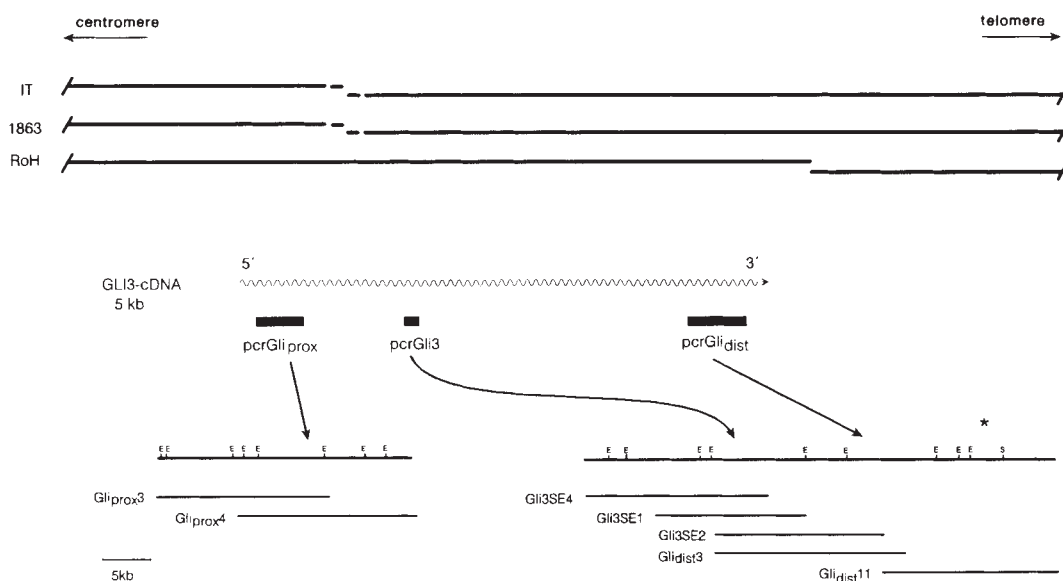


FIG. 2 A physical map of the GCPS/*GLI3* region. The translocation breakpoints of the three GCPS patients are shown on the top in relation to the *GLI3* cDNA sequence. The position of the PCR probes used is indicated by black boxes. A restriction map derived from genomic phage clones isolated by screening with the PCR probes or phage end fragments is drawn on a larger scale (E, *EcoRI*; S, *SalI*). The location of the RoH breakpoint spanned by

Gli3dist11 is indicated by an asterisk.

METHODS. Overlapping phage clones were isolated from a genomic *MboI* partial digest library in EMBL3 using the PCR probes described in Fig. 1. Clone *Gli3dist11* was isolated by chromosome walking with the 6-kp *EcoRI/SalI* end-fragment of phage *Gli3dist3* subcloned into pBluescript II SK- (Stratagene). All subcloning and hybridization was according to standard procedures¹⁸.

gene is expressed in tissues such as testes, myometrium, placenta, colon and lung as an 8.5-kilobase messenger RNA and the protein product (relative molecular mass, 190,000) shows sequence-specific DNA binding^{5,6}. On our panel this probe, pcrGli3, mapped distal to the Greig translocation breakpoints in patients 1863 (46, XY, t(3; 7) (p21.1; p13)) and IT (46, XX, t(6; 7)(q12; p13)), but proximal to the one in RoH (46, XX, t(6; 7)(q27; p13)) (Fig. 1d). Thus, *GLI3* maps very closely or may even be identical to the GCPS gene.

To define the position of the three translocation breakpoints along the *GLI3* gene we used pcrGli_{prox} and pcrGli_{dist}, sequences corresponding to proximal and distal segments of the *GLI3* complementary DNA sequence⁶ (Fig. 2), amplified by PCR from lung cDNA. PcrGli_{prox} maps proximal to all GCPS translocation breakpoints (Fig. 1c), indicating that two of the translocations split the *GLI3* gene within the first third of the coding region, thereby precluding the formation of a normal functioning protein. This also establishes the transcriptional orientation of *GLI3* from centromere to telomere. Probes from a 30-kb genomic phage contig selected by pcrGli_{prox} did not detect the translocation breakpoints, suggesting that the 5' part of this gene is spread over a large genomic distance (Fig. 2).

On the other hand, pcrGli_{dist}, like pcrGli3, maps proximal to the RoH breakpoint and distal to the 1863 and IT breakpoints (Fig. 1e). The third translocation breakpoint thus has to be downstream of the 3' end of *GLI3* (Fig. 2). Analysis of five independent cDNA clones isolated from a human fetal kidney library¹⁰ using pcrGli_{dist} as probe showed no evidence for alternatively spliced transcripts in this region, which might extend to the RoH translocation breakpoint. Thus, the translocation probably falls distal to the last exon of *GLI3*.

To map the distal translocation breakpoint, we isolated a 50-kb phage contig, including 15 kb of DNA downstream of the *GLI3* coding sequence. The RoH translocation breakpoint was found by hybrid mapping in a 3.5-kb *EcoRI/SalI* fragment about 10 kb downstream of *GLI3* (Figs 1f, and 2). In this case a *cis*-acting element might have been brought into the neighbourhood of *GLI3*, thereby modifying its expression. The influence of 3'-acting domains has been well documented in variant Burkitt

lymphomas, in which translocations of more than 10 kb downstream of the *c-myc* gene deregulate its expression¹¹.

Our results support the assumption that mutations affecting only one allele of the zinc finger gene *GLI3* cause the GCPS syndrome. With the potential function as a transcriptional regulator, *GLI3* could well control limb and craniofacial development, perhaps through involvement in the generation of a morphogen gradient, as proposed for vertebrate limb formation¹². The fact that *GLI3* is widely expressed in adult tissues, with only some of these being phenotypically altered by heterozygous mutation, implies that precise control of *GLI3* expression is crucial only at certain developmental stages. Studies of mouse mutant extra toes (Xt)¹³, and anterior digit deformity (add)¹⁴, which are presumed to have similarities with GCPS, may reveal more about the function of *GLI3*. Various finger proteins, among them those from the *GLI-Krüppel* gene family, have key roles in *Drosophila* and *Xenopus* development. Our results provide an example of a zinc-finger gene that is mutated in a hereditary human malformation. □

Received 23 April; accepted 21 June 1991.

- Greig, D. M. *Edin. med. J.* **33**, 189-218 (1926).
- Gollop, T. R. & Fontes, L. R. *Am. J. med. Genet.* **22**, 59-68 (1985).
- Tommerup, N. & Nielsen, F. *Am. J. med. Genet.* **16**, 313-321 (1983).
- Krüger, G. et al. *Am. J. med. Genet.* **32**, 411-416 (1989).
- Ruppert, J. M. et al. *Molec. cell. Biol.* **8**, 3104-3113 (1988).
- Ruppert, J. M., Vogelstein, B., Arheden, K. & Kinzler, K. W. *Molec. cell. Biol.* **10**, 5408-5415 (1990).
- Wagner, K., Kroisel, P. M. & Rosenkranz, W. *Genomics* **8**, 487-491 (1990).
- Brueton, L., Huson, S. M., Winter, R. M. & Williamson, R. *Am. J. med. Genet.* **31**, 799-804 (1988).
- Drabkin, H. et al. *Genomics* **4**, 518-529 (1989).
- Gessler, M. et al. *Nature* **343**, 774-778 (1990).
- Bornkamm, G. W., Pollack, A. & Eick, D. in *Cellular Oncogene Activation* (ed. Klein, G.) 223-273 (Dekker, New York, 1988).
- Eichele, G. *Trends Genet.* **5**, 246-251 (1989).
- Winter, R. & Huson, M. *Am. J. med. Genet.* **31**, 793-798 (1988).
- Pohl, T. M., Mattei, M.-G. & Rüther, U. *Development* **110**, 1153-1157 (1990).
- Rommens, J. M. et al. *Am. J. hum. Genet.* **43**, 645-633 (1988).
- Jobs, A. et al. *Hum. Genet.* **84**, 147-150 (1990).
- Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Vortkamp, A. et al. *Genomics* (in the press).

ACKNOWLEDGEMENTS. We thank N. Tommerup and L. Pelz for fibroblast cell lines 1863 and RoH. C. Croce for hybrid 5387-3cl10 and S. Orkin for an aliquot of the EMBL3 library. This work was supported by a grant from the Deutsche Forschungsgemeinschaft.

Amino-terminal domains of *c-myc* and *N-myc* proteins mediate binding to the retinoblastoma gene product

Anil K. Rustgi[‡], Nicholas Dyson[†] & Rene Bernards^{*}

Divisions of^{*} Molecular Genetics and [†] Molecular Oncology, The Cancer Center of the Massachusetts General Hospital and Harvard Medical School, and [‡]Gastrointestinal Unit of Massachusetts General Hospital, 149 13th Street, Charlestown, Massachusetts, USA

THE proteins encoded by the *myc* gene family are involved in the control of cell proliferation and differentiation, and aberrant expression of *myc* proteins has been implicated in the genesis of a variety of neoplasms¹. In the carboxyl terminus, *myc* proteins have two domains that encode a basic domain/helix-loop-helix and a leucine zipper motif, respectively. These motifs are involved both in DNA binding and in protein dimerization²⁻⁵. In addition, *myc* protein family members share several regions of highly conserved amino acids in their amino termini that are essential for transformation^{6,7}. We report here that an N-terminal domain present in both the *c-myc* and *N-myc* proteins mediates binding to the retinoblastoma gene product, pRb. We show that the human papilloma virus E7 protein competes with *c-myc* for binding to pRb, indicating that these proteins share overlapping binding sites on pRb. Furthermore, a mutant Rb protein from a human tumour cell line that carried a 35-amino-acid deletion in its C terminus failed to bind to *c-myc*. Our results suggest that *c-myc* and pRb cooperate through direct binding to control cell proliferation.

The study of proteins that interact with *c-myc* has been hindered by the inherent insolubility of *myc* proteins. Conditions that allow extraction of *c-myc* from the nucleus are too stringent to preserve protein-protein interactions, whereas mild extraction conditions fail to solubilize *c-myc*⁸. To circumvent this problem, we used a bacterial expression vector in which the glutathione S-transferase (GST) gene is linked to a DNA fragment that encodes the first 204 amino acids of human *c-myc*. The encoded GSTΔ*myc* fusion protein is highly soluble and can be readily purified from bacterial lysates by binding to glutathione agarose beads (ref. 9, and D. Hill, unpublished observations).

To investigate whether cellular proteins form specific complexes with any of the conserved N-terminal domains of the *c-myc* protein, we labelled various human cell lines with ³⁵S-methionine and prepared mild-detergent lysates from labelled cells. These extracts were then incubated with either GST protein bound to agarose beads (GST beads) or with GST-*myc* chimaeric protein bound to agarose beads (GSTΔ*myc* beads) as described¹⁰. Cellular proteins bound by both types of beads were analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). These experiments indicated that a small number of cellular proteins were bound by GSTΔ*myc* beads, but not by control GST beads (A.K.R. and R.B., unpublished data). One of the cellular proteins that was specifically bound by GSTΔ*myc* beads had a relative molecular mass of about 105,000 (*M_r*, 105K).

As pRb has a similar *M_r*¹¹, we tested whether GSTΔ*myc* could form a specific complex with pRb. We incubated GSTΔ*myc* beads with unlabelled mild-detergent lysates from ML1 leukaemia cells (which express wild-type Rb protein) and separated bound proteins by SDS-PAGE. The presence of pRb was detected by western blot analysis using monoclonal antibodies to pRb. As a positive control we used GST-E1A beads, as

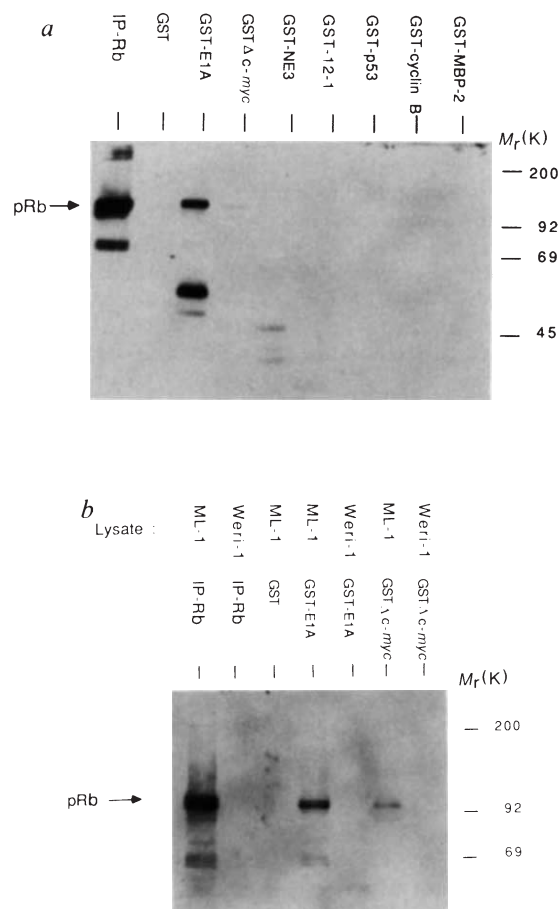


FIG. 1 Western blot analysis of Rb protein bound by GST fusion protein beads. Mild-detergent extracts were prepared from ML1 leukaemia cells or WERI-1 retinoblastoma cells as described¹² and incubated with agarose beads loaded with 20 μg GST fusion protein. As a control, pRb protein was immunoprecipitated from cell lysates for comparison (lanes labelled IP-Rb). Bound cellular proteins were released by boiling in SDS buffer and separated on a 6% acrylamide-SDS gel. Proteins were then transferred to nitrocellulose. The pRb protein was detected with monoclonal antibodies to pRb and an alkaline phosphatase-conjugated second antibody. *a*, Proteins bound using extracts from 2×10^7 ML1 leukaemia cells per lane. *b*, Proteins bound using both ML1 extracts and extracts derived from an equal number of WERI-1 retinoblastoma cells. The following GST fusion proteins were used: GST alone (lanes labelled GST), GST fused to the first 204 amino acids of human *c-myc* protein (GSTΔ*myc*), GST fused to the Ad5 243 amino acid E1A protein (GST-E1A), GST fused to the N-*myc* basic domain/helix-loop-helix region (GST-NE3), GST fused to the N-*myc* helix-loop-helix region (GST-12-1), GST fused to human p53 carrying a mutation in codon 273 (GST-p53), GST fused to cyclin B (GST-cyclin B) and GST fused to the MBP-2 transcription factor gene (ref. 26, GST-MBP-2). The lower-*M_r* protein species seen in the western blots are probably the result both of pRb protein degradation and of cross-reactivity of the pRb antisera with the bacterial GST fusion proteins.

METHODS. GST fusion proteins were prepared basically as described by Kaelin *et al.*¹⁰ with the modification that GST fusion proteins bound to glutathione-agarose beads were eluted with 5 mM glutathione. After elution, fusion proteins were dialysed overnight in 0.1 M sodium phosphate buffer, pH 7.8, and protein concentration was measured by modified Bradford assay. Dialysed proteins were checked for integrity by SDS-PAGE followed by Coomassie blue staining. This procedure of protein elution followed by rebinding to glutathione-agarose beads makes it possible to check protein integrity and quantify the amount of GST fusion protein loaded on the beads. To generate GST fusion protein-loaded beads, 50 μl glutathione-agarose beads (50% slurry, Sigma) was incubated with 20 μg GST fusion protein for 30 min at 4 °C. After this, beads were washed three times with 1 ml ELB buffer¹². Beads were then incubated with cell lysates for 2 h at 4 °C and washed five times 1 ml ELB buffer. Bound proteins were released by boiling in SDS buffer and separated on a polyacrylamide-SDS gel. Western blotting was as described²⁷.

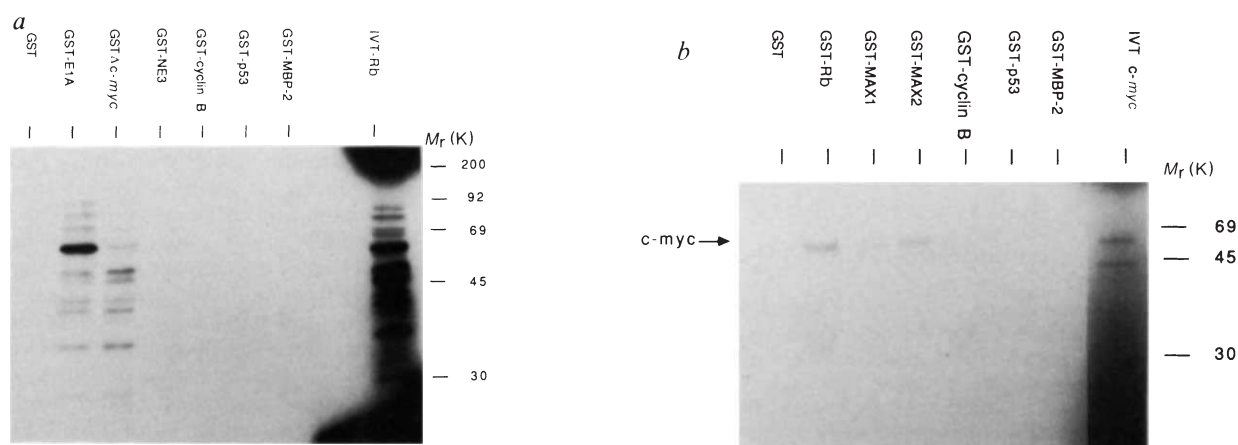


FIG. 2 Binding of *in vitro*-translated proteins to GST fusion protein beads. ^{35}S -labelled *in vitro*-translated human pRb and human *c-myc* proteins were incubated with agarose beads loaded with GST fusion proteins as indicated. For comparison, 1 μl total *in vitro*-translated protein was loaded on the gel (lanes labelled IVT-Rb and IVT *c-myc*). The 45K protein species seen in the *c-myc in vitro* translation is a nonspecific protein also found in unprogrammed rabbit reticulocyte lysates. Bound proteins were separated on a 10% polyacrylamide gel and detected by fluorography. *a*, *In vitro*-translated pRb protein bound to GST fusion protein beads. *b*, *In vitro*-translated *c-myc* protein bound to GST fusion protein beads. *c*, *In vitro*-translated murine *N-myc* protein bound to GST fusion protein beads.

METHODS. GST fusion protein (1 μg) was loaded on 20 μl glutathione-agarose beads as described in the legend to Fig. 1. *In vitro*-translated proteins (10 μl) were mixed with fusion protein beads in 0.5 ml NETN buffer¹⁰ and allowed to bind for 2 h at 4 °C. After this, beads were washed four times with 1 ml NETN. Bound proteins were released by boiling in SDS buffer and separated on a 10% acrylamide gel. The pRb and *c-myc* proteins were generated by sequential *in vitro* transcription and translation (Stratagene). GST fusion proteins used were as described in Fig. 1. GST-MAX1 and GST-MAX-2 represent GST fused to MAX, a *c-myc* binding protein. MAX-1 and MAX2 differ by nine amino acids as a result of alternative splicing of the MAX transcript³.

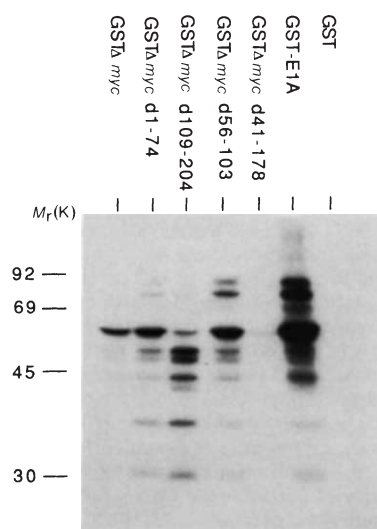


FIG. 3 Binding of pRb to mutant *c-myc* proteins. Agarose beads harbouring GST Δ *myc* protein or GST Δ *myc* mutant proteins were incubated with ^{35}S -methionine labelled *in vitro* translated pRb as described in the legend to Fig. 2. Bound pRb protein was separated on a 12% polyacrylamide gel and detected by fluorography. GST Δ *myc*d1-74 contains amino acids 75-204 of human *c-myc* linked to GST; GST Δ *myc*d109-204 contains amino acids 1-108 of *c-myc*; GST Δ *myc*d56-103 contains amino acids 1-55 and 103-204 of *c-myc*; GST Δ *myc*d41-178 has amino acids 1-41 and 178-204 of *c-myc*. The *myc* deletion mutants were generated by polymerase chain reaction amplification with specific *c-myc* primers using either wild-type *c-myc* plasmids or mutant *c-myc* plasmids⁶ as a template.

adenovirus E1A-encoded proteins bind strongly to pRb (ref. 12). Both GST-E1A beads and GST Δ *myc* beads bind a 105K protein that is recognized by pRb monoclonal antibodies (Fig. 1a). This protein comigrated with pRb immunoprecipitated with pRb antibodies from the same cell lysates (lane labelled IP-Rb) and was not bound by GST beads or by five other randomly chosen GST fusion proteins (Fig. 1a). To confirm that the 105K protein bound by GST-E1A and GST Δ *myc* beads was pRb, we repeated this experiment with lysates from WERI-1 retinoblastoma cells which have a homozygous deletion of the *Rb* gene¹³. The 105K protein was not bound by GST-E1A or GST Δ *myc* beads in WERI-1 lysates (Fig. 1b). Taken together, these results indicate that GST Δ *myc* protein can form a specific complex with pRb *in vitro*, although, as expected for a cellular Rb binding protein, the affinity does not seem to be as high as that of GST-E1A for pRb (Fig. 1a, b).

To study further the association between *c-myc* and pRb, we generated ^{35}S -methionine-labelled pRb by *in vitro* transcription and translation. This yields a characteristic series of Rb polypeptides that share a common C terminus generated by initiation of translation at internal start codons^{12,14}. *In vitro* synthesized pRb was incubated with GST-E1A beads, GST Δ *myc* beads or a series of control GST fusion protein beads and bound proteins were separated by SDS-PAGE. GST-E1A and GST Δ *myc* beads, but not control GST-protein beads, bound to *in vitro*-translated pRb (Fig. 2a).

We then tested whether *in vitro*-translated full-length human *c-myc* protein could form a specific complex with a GST-Rb fusion protein that has only the pRb domain required for binding to E1A (amino acids 379-792 of pRb; ref. 10). As a positive control, we used GST-MAX beads, as the protein MAX forms

heterodimers with *c-myc* through the helix-loop-helix/leucine zipper motifs³. Human *c-myc* translated *in vitro* bound to GST-Rb and GST-MAX beads with equal affinity (Fig. 2b). No binding of *c-myc* was seen to four other GST fusion proteins, including another tumour suppressor gene product, p53.

The N-terminal half of *c-myc* protein has several regions of high sequence homology to other *myc* protein family members⁷. We therefore tested whether *in vitro*-translated N-*myc* could bind to GST-Rb beads. Murine N-*myc* translated *in vitro* binds to GST-RB beads and GST-MAX beads, but failed to bind to other GST fusion proteins (Fig. 2c). Our data seem to indicate that *c-myc* binds pRb better than N-*myc* (Fig. 2b, c). But because human *c-myc* and murine N-*myc* might differ in their affinity for MAX, this conclusion is not warranted by our data. We nonetheless conclude that the ability to bind pRb may be a feature common to *myc* protein family members. It is likely that pRb binding requires one or more of the highly conserved *myc* homology domains.

To map the domains in the *c-myc* protein involved in binding to pRb, we used a series of GST-*myc* mutant proteins. These mutant *myc* proteins were used in a binding assay with *in*

vitro-translated ³⁵S-methionine-labelled pRb protein. Both deletion of amino acids 1-74 and 109-204 of *c-myc* protein yielded proteins that bound to pRb (Fig. 3). Deletion of amino acids 56-103 also did not affect binding to pRb. But very weak binding was observed when amino acids 41-178 were deleted in *c-myc*. Thus at least two distinct regions in the *c-myc* protein mediate binding to pRb.

The Rb protein species generated by *in vitro* translation from internal AUG start codons represent a series of N-terminal truncations of pRb that can be used to map roughly the binding site of the *c-myc* protein on pRb¹⁴. High-affinity binding of E1A to pRb requires two domains of the Rb protein comprising amino acids 393-572 and 646-772, respectively¹⁴. The *in vitro* translation products of pRb with $M_r \geq 56$ K will therefore contain both E1A binding domains. Under the experimental conditions used, GST-E1A also shows a preference for ≥ 56 K pRb products, but additionally can bind to pRb *in vitro* translation products of ≥ 28 K with reduced affinity. GST-E1A protein did not bind to pRb *in vitro* translation products smaller than 28K, indicating that a region between amino acids 695 and 761 is essential for binding of GST-E1A to pRb. Similarly, GST Δ *myc*

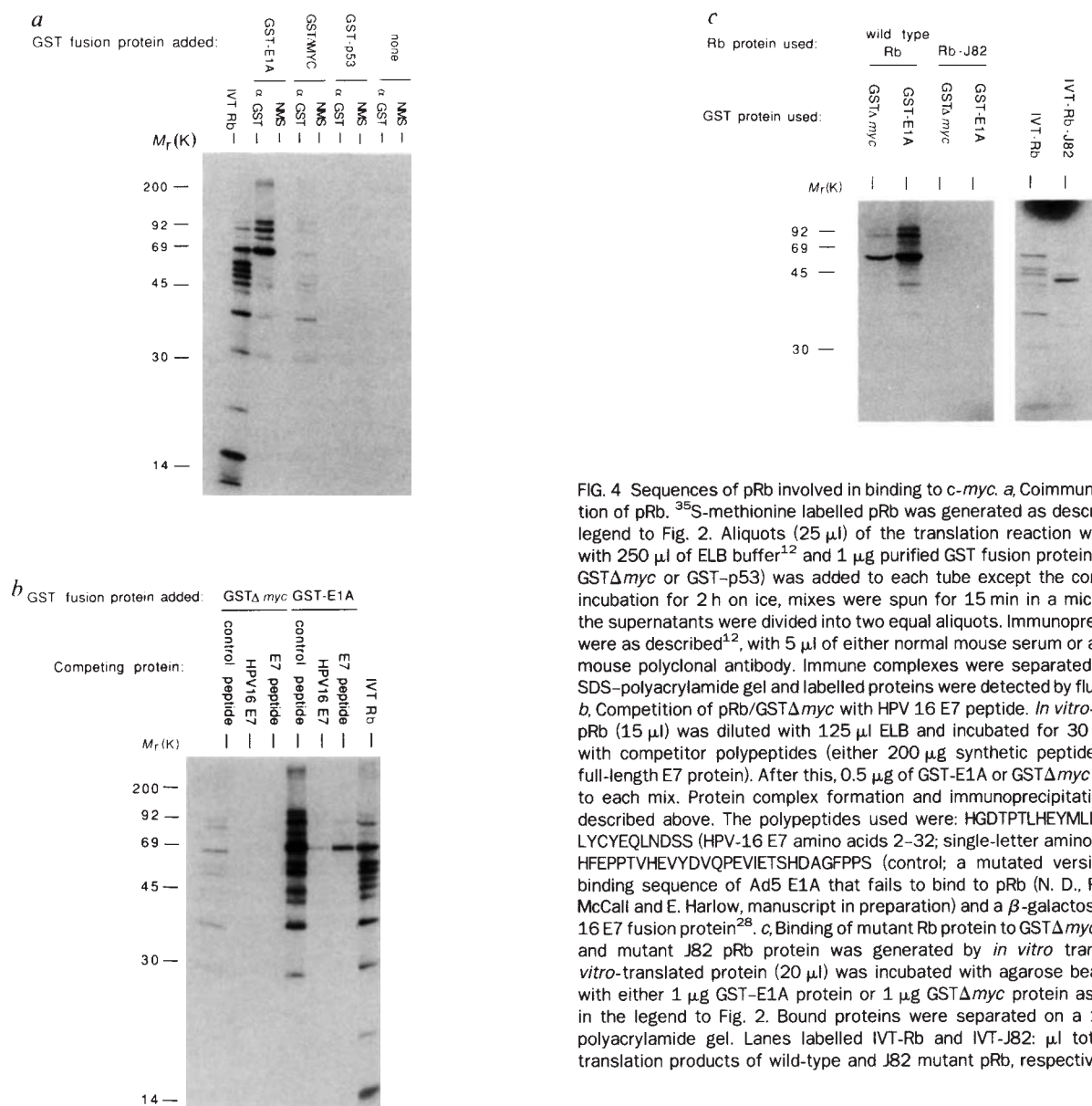


FIG. 4 Sequences of pRb involved in binding to *c-myc*. **a**, Coimmunoprecipitation of pRb. ³⁵S-methionine labelled pRb was generated as described in the legend to Fig. 2. Aliquots (25 μ l) of the translation reaction were diluted with 250 μ l of ELB buffer¹² and 1 μ g purified GST fusion protein (GST-E1A, GST Δ *myc* or GST-p53) was added to each tube except the control. After incubation for 2 h on ice, mixes were spun for 15 min in a microfuge and the supernatants were divided into two equal aliquots. Immunoprecipitations were as described¹², with 5 μ l of either normal mouse serum or an anti-GST mouse polyclonal antibody. Immune complexes were separated on a 12% SDS-polyacrylamide gel and labelled proteins were detected by fluorography. **b**, Competition of pRb/GST Δ *myc* with HPV 16 E7 peptide. *In vitro*-translated pRb (15 μ l) was diluted with 125 μ l ELB and incubated for 30 min on ice with competitor polypeptides (either 200 μ g synthetic peptide or 25 μ g full-length E7 protein). After this, 0.5 μ g of GST-E1A or GST Δ *myc* was added to each mix. Protein complex formation and immunoprecipitation was as described above. The polypeptides used were: HGDPTLHEYMLDLQPETTD-LYCYEQLNDSS (HPV-16 E7 amino acids 2-32; single-letter amino-acid code) HFEPPTVHEVYDVQPEVIETSHDAGFPSPS (control; a mutated version of pRb binding sequence of Ad5 E1A that fails to bind to pRb (N. D., P. Guida, C. McCall and E. Harlow, manuscript in preparation) and a β -galactosidase-HPV 16 E7 fusion protein²⁸. **c**, Binding of mutant Rb protein to GST Δ *myc*. Wild-type and mutant J82 pRb protein was generated by *in vitro* translation. *In vitro*-translated protein (20 μ l) was incubated with agarose beads loaded with either 1 μ g GST-E1A protein or 1 μ g GST Δ *myc* protein as described in the legend to Fig. 2. Bound proteins were separated on a 12% SDS-polyacrylamide gel. Lanes labelled IVT-Rb and IVT-J82: μ l total *in vitro* translation products of wild-type and J82 mutant pRb, respectively.

protein bound to pRb translation products $\geq 28K$, but did not show increased affinity for translation products $\geq 56K$, as seen with GST-E1A (Fig. 4a). Furthermore, no binding to GST-E1A and GST Δ myc protein was observed with an *in vitro*-translated pRb polypeptide that contained a C terminus at amino acid 667 of the wild-type protein (data not shown). These data suggest that the binding sites for E1A and c-myc protein pRb may include similar regions in the C terminus of pRb.

To investigate further whether c-myc and DNA tumour virus-encoded proteins bind to similar sites on pRb, we did a competition experiment. We used human papilloma virus 16 (HPV-16) E7 protein for this experiment as this protein binds strongly to pRb¹⁵. Bacterially synthesized HPV-16 E7 protein or an HPV-16 E7 peptide that contained the pRb binding site on the E7 protein were incubated with *in vitro*-translated pRb. After this, GST Δ myc protein or GST-E1A protein was added and binding of pRb to GST fusion proteins was assayed. Both HPV-16 E7 protein and HPV-16 E7 peptide, but not control peptide, competed efficiently for binding of both GST Δ myc and GST-E1A protein to pRb (Fig. 4b). We conclude that c-myc, HPV-16 E7 and adenovirus E1A proteins have overlapping binding sites on pRb.

The C-terminal region of pRb to which c-myc binds is frequently deleted in human cancer¹⁶⁻¹⁸. This raised the possibility that in these tumour cells the interaction between c-myc protein and pRb is perturbed. To test this hypothesis, we used a cloned pRb cDNA from the J82 bladder carcinoma cell line which carries a 35-amino-acid C-terminal deletion in pRb (amino acids 697-731; ref. 16). *In vitro*-translated wild-type pRb and mutant J82 pRb was subsequently tested for binding to GST Δ myc. The mutant pRb from the J82 bladder carcinoma binds both GST-E1A and GST Δ myc with greatly reduced affinity (Fig. 4c). We conclude that the 35-amino-acid deletion of pRb in the J82 carcinoma cell line greatly alters the ability of pRb to bind to c-myc protein.

Interesting parallels exist between the c-myc and the adenovirus E1A genes. Both encode nuclear phosphoproteins that cooperate with an activated *ras* oncogene in transformation^{19,20}. Additionally, *myc* and E1A are structurally distantly related^{21,22}. We now extend these similarities by showing that c-myc protein, like E1A, can bind to pRb. Because full-length c-myc protein is highly insoluble, it was necessary to study the interaction between c-myc protein and pRb *in vitro* with a truncated soluble form of c-myc protein. Our findings are nevertheless intriguing because the region of pRb to which c-myc protein binds is frequently deleted in human tumours¹⁶⁻¹⁸. This suggests that c-myc protein interacts with a physiologically important site of pRb.

The ability of viral oncogenes to transform is intimately linked to their ability to bind pRb²³⁻²⁴. Our finding that the HPV-16 E7-encoded protein competes with c-myc protein for binding to pRb suggests that both proteins control cell proliferation by binding to overlapping sites on pRb. In this regard it is noteworthy that either antisense c-myc oligonucleotides or pRb overexpression results in a block in cell cycle progression from G1 to S phase (ref. 25, and S. Friend, personal communication). On the basis of our observations, it is tempting to speculate that dominant and recessive oncogenes can cooperate through direct binding to control progression through the cell cycle. □

12. Whyte, P. *et al.* *Nature* **334**, 124-129 (1988).
13. Friend, S. H. *et al.* *Nature* **323**, 643-646 (1986).
14. Hu, Q., Dyson, N. & Harlow, E. *EMBO J.* **9**, 1147-1155 (1990).
15. Dyson, N., Howley, P. M., Münger, K. & Harlow, E. *Science* **243**, 934-936 (1989).
16. Horowitz, J. M. *et al.* *Science* **243**, 937-940 (1989).
17. Yandell, D. W. *et al.* *New Engl. J. Med.* **321**, 1689-1695 (1989).
18. Dunn, J. M. *et al.* *Molec. cell. Biol.* **9**, 4596-4604 (1989).
19. Land, H., Parada, L. F. & Weinberg, R. A. *Nature* **304**, 596-602 (1983).
20. Ruley, H. E. *Nature* **304**, 602-606 (1983).
21. Ralston, R. & Bishop, J. M. *Nature* **306**, 803-806 (1983).
22. Figge, J., Wester, T., Smith, T. F. & Paucha, E. *J. Virol.* **62**, 1814-1818 (1988).
23. Whyte, P., Williamson, N. M. & Harlow, E. *Cell* **56**, 67-75 (1989).
24. De Caprio, J. A. *et al.* *Cell* **54**, 275-283 (1988).
25. Heikkilä, R. *et al.* *Nature* **328**, 445-449 (1987).
26. Rustgi, A. K., Van't Veer, L. J. & Bernards, R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8707-8710 (1990).
27. Windle, J. J. *et al.* *Nature* **343**, 665-669 (1990).
28. Jones, R. E. *et al.* *J. biol. Chem.* **265**, 12782-12785 (1990).

ACKNOWLEDGEMENTS. We thank M. Billaud, A. Fattaey, T. Frebourg, D. Hill, W. Kaelin, S. Shiff, L.-H. Tsai and L. van't Veer for GST fusion proteins, D. Patrick for HPV E7 protein, R. Kingston for the c-myc *in vitro* expression vector, J. Stone for mutant c-myc plasmids, J. Horowitz for the pRb-J82 plasmid, E. Harlow for discussions and critical reading of the manuscript, and M. Dembski for expert technical assistance. This work was supported by grants from the National Institutes of Health and the Searle Scholarship Foundation (R.B.), and an Institutional Grant from the Massachusetts General Hospital (N.D.). A.K.R. was supported by the American Cancer Society and by a Clinical Investigator Award.

Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*

Thuan Q. Trinh* & Richard R. Sinden†

Department of Molecular Genetics, Biochemistry and Microbiology, University of Cincinnati College of Medicine, 231 Bethesda Avenue, Cincinnati, Ohio 45267-0524, USA

WHEN present in single-stranded DNA, palindromic or quasi-palindromic sequences have the potential to form complex secondary structures, including hairpins, which may facilitate inter-strand misalignment of direct repeats and be responsible for diverse types of replication-based mutations, including deletions, additions, frameshifts and duplications¹⁻⁵. In regions of palindromic symmetry, specific deletion events may involve the formation of a hairpin or other DNA secondary structures which can stabilize the misalignment of direct repeats^{1,2}. One model suggests that these deletions occur during DNA replication by slippage of the template strand and misalignment with the progeny strand^{6,7}. The concurrent DNA replication model, involving an asymmetric dimeric DNA polymerase III complex which replicates the leading and lagging strands⁸, has significant implications for mutagenesis. The intermittent looping of the lagging strand template, and the fact that the lagging strand template may contain a region of single-stranded DNA the length of an Okazaki fragment, provides an opportunity for DNA secondary-structure formation and misalignment. Here we report our design of a palindromic fragment to create an 'asymmetric palindromic insert' in the chloramphenicol acetyltransferase gene of plasmid pBR325. The frequency with which the insert was deleted in *Escherichia coli* depends on the orientation of the gene in the plasmid. Our results suggest that replication-dependent deletion between direct repeats may occur preferentially in the lagging strand.

Plasmid pBR325, a ColE1-derived plasmid, has a unidirectional origin of replication and requires many host-encoded proteins for replication⁹. The transcribed strand of the chloramphenicol acetyltransferase (*CAT*) gene is the leading template strand and the non-coding strand the lagging strand. Reversing the direction of the *CAT* gene changes the relationship between the coding strand and the leading and lagging

Received 22 May; accepted 15 July 1991.

1. Lüscher, B. & Eisenman, R. N. *Genes Dev.* **4**, 2025-2035 (1990).
2. Murre, C. M. M., McCaw, P. S. & Baltimore, D. *Cell* **56**, 777-783 (1989).
3. Blackwood, E. M. & Eisenman, R. N. *Science* **251**, 1211-1217 (1991).
4. Blackwell, T. K. *et al.* *Science* **250**, 1149-1151 (1990).
5. Pendergast, G. C. & Ziff, E. B. *Science* **251**, 186-189 (1991).
6. Stone, J. *et al.* *Molec. cell. Biol.* **7**, 1697-1709 (1987).
7. Sugiyama, A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 9144-9148 (1989).
8. Moore, J. P., Hancock, D. C., Littlewood, T. D. & Evan, G. I. *Oncogene Res.* **2**, 65-80 (1987).
9. Smith, D. B. & Johnson, K. S. *Gene* **67**, 31-40 (1988).
10. Kaelin, W. G. *et al.* *Cell* **64**, 521-532 (1991).
11. Lee, W. H. *et al.* *Nature* **329**, 642-645 (1987).

* To whom correspondence should be addressed.

† Present address: Merck, Sharp, and Dohme Research Laboratories, Department of Membrane Biochemistry and Biophysics PO Box 2000, Rahway, New Jersey 07065, USA.

Plasmids	Nucleotide sequence at the <i>EcoRI</i> site.
pBR325p1	<pre> <----- <----- <----- gctcatccggAATTCCTATAGGAATTCaattccgtagtgca ----->----->----- </pre>
pBR523p1	<pre> <----- <----- <----- tgccatacggAATTAGAAATTCCTATAGGaattccggatgagc ----->----->----- </pre>
pBR325n1	<pre> <----- <----- gctcatccggAATTCGTCTGATGCACGaattccgtagtgca ----->----->----- </pre>
pBR523n1	<pre> <----- <----- tgccatacggAATTCGTGCATCAGACGaattccggatgagc ----->----->----- </pre>

FIG. 1 Nucleotide sequence of the plasmids around the *EcoRI* site in the *CAT* gene of pBR325 and pBR523. Shown are the sequences of the 'deletion insert' (capital letters) and the surrounding *CAT* gene sequence (lower-case). Inserts were chemically synthesized as *EcoRI* fragments and cloned into pBR325. To construct the pBR523 derivatives, the *CAT* gene was reversed by cutting with *AsuII* and religating. A difference in the *PvuII* restriction patterns of pBR325 and pBR523 derivatives was used to screen potential pBR523 clones. The nucleotide sequence of all plasmids was confirmed by DNA sequencing. The various direct repeats are indicated by arrows above the sequence pointing in the direction of replication of the leading strand. The inverted repeat in the p1 insert is indicated by arrows below the sequence.

template strands of replication. Reversing the gene, rather than reversing the insert within the *CAT* gene, preserves the relationship between the insert and flanking DNA sequence that might influence the frequency of mutation.

The frequency of spontaneous deletion between direct repeats represents the sum of all possible events that result from homologous or illegitimate recombination, DNA repair activities or errors of DNA replication. If deletion between direct repeats occurs with equal frequency in the leading and lagging strands, there would be no effect of reversing the *CAT* gene. During replication, if deletion occurs preferentially in either the leading or lagging strand, deletion might occur at different frequencies. For a 'symmetric deletion insert', the mechanism of deletion or the DNA secondary structures that may be involved in deletion will be the same in either the leading or lagging

strand. Consequently, the frequency of deletion should be unchanged by reversing the *CAT* gene. But for an 'asymmetric deletion insert', the DNA secondary structures that may be involved in deletion will be different in the leading and lagging strands, and thus a different deletion frequency will be observed when the *CAT* gene is reversed.

We have designed several 17- and 18-base-pair symmetric and asymmetric deletion inserts and have shown that the frequency of deletion from the *EcoRI* site in the *CAT* gene in pBR325 can vary over more than a 100-fold range in *recA*⁻ *E. coli*. Reversion to chloramphenicol resistance (Cm^r) occurs primarily by deletion between the direct repeats, which include the *EcoRI* sites, resulting in a restoration of the *CAT* gene (T.Q.T. and R.R.S., manuscript in preparation). Here we examine the effect of reversing the *CAT* gene on the frequency of deletion of a

A pBR325n1, Leading Strand

a

```

      <----- <-----
5' gctcatccggAATTCGTCTGATGCACGaattccgtagtgca 3'
3' cgagtaggccttaaGCAGACTACGTGCTTAaggcataccgt 5'

```

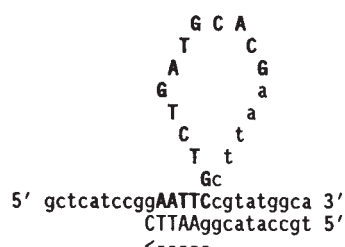
b

```

      <----- <-----
5' gctcatccggAATTCGTCTGATGCACGaattccgtagtgca 3'
3'          CTAAaggcataccgt 5'
      ----->----->

```

C



B pBR325n1, Lagging Strand

a

```

      -----> ----->
5' gctcatccggAATTCGTCTGATGCACGaattccgtagtgca 3'
3' cgagtaggccttaaGCAGACTACGTGCTTAaggcataccgt 5'

```

b

```

      ----->
5' gctcatccggAATTC
3' cgagtaggccttaaGCAGACTACGTGCTTAaggcataccgt 5'

```

C

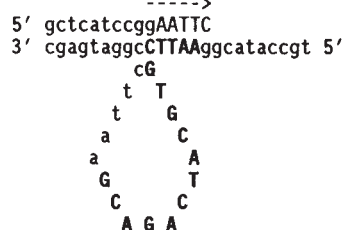


FIG. 2 Potential structural intermediates involved in deletion of a non-palindromic symmetric deletion insert. A, Potential DNA secondary structure in the leading strand of pBR325n1 containing a 'symmetric non-palindromic insert'. a, Nucleotide sequence of the 17-bp n1 insert and flanking DNA. The n1 insert is in upper-case bold characters. The direct repeats are represented by arrows above the sequence. b, DNA replication of the first copy of the

direct repeat 5'GAATTC3'. c, The template strand slips and the second (left-most) copy of the direct repeat in the template pairs with the direct repeat in the progeny strand. Replication continues resulting in deletion of the fragment between direct repeats in the progeny strand. B, Potential DNA secondary structure in the lagging strand of pBR325n1. Events shown are analogous to that described above for the leading strand.

symmetric (n1) and asymmetric (p1) insert (Fig. 1). These pBR325-based plasmids reverted to a Cm^r phenotype at frequencies of $2\text{--}44 \times 10^{-9}$ Cm^r revertants per viable cell in the *recA*⁻ *E. coli* strain DH5 (Table 1). Reversing the *CAT* gene in the pBR523-based plasmid did not alter the Cm^r reversion frequency of the symmetric n1 deletion insert. The Cm^r reversion frequencies of two other symmetric deletion inserts (one non-palindromic, n2, and one palindromic, p1) were also similar when the gene was reversed (T.Q.T. and R.R.S., manuscript in preparation). When the *CAT* gene was reversed, the reversion frequency for the asymmetric p1 deletion insert was reduced by about a factor of 20. This suggests that deletion occurs preferentially in either the leading or lagging strand.

The deletion inserts in the *CAT* gene are symmetric or asymmetric with respect to replication (Fig. 1) and the potential for misalignment that can lead to deletion (Figs 2, 3). The n1 deletion insert represents a symmetric situation because mispaired intermediates containing an unpaired loop can form during replication in the leading and lagging strands (Fig. 2). The p1 insert is asymmetric with respect to replication of the direct-repeated and inverted-repeated elements (Fig. 1). As a con-

sequence of this asymmetry, different DNA secondary structures may be involved in deletion in the leading and lagging strands, as explained in Fig. 3. In the leading strand of pBR325p1, misalignment may involve the formation of a loop in the DNA between the direct repeats (A). In the lagging strand, a hairpin may form, stabilizing the misalignment of direct repeats and leading to deletion of the insert (B). The secondary structures which may occur at the 3' end of the leading strand of pBR325p1 are those that could occur at the 3' end of the lagging strand of pBR523p1. Similarly, events that could occur at the 3' end of the lagging strand of pBR325p1 are those that could also occur at the 3' end of the leading strand of pBR523p1.

Reversing the *CAT* gene in pBR325p1 decreased the Cm^r reversion frequency by a factor of 20 (Table 1). The same result has been obtained in two other *recA*⁻ strains HB101 and JTT1. This result is consistent with the frequency of mutagenesis differing between the leading and the lagging strands. In the lagging strand of pBR325p1, where misalignment may be stabilized by a hairpin, a deletion frequency of 44×10^{-9} was observed. In cases of n1 and the lagging strand of pBR523p1, where deletion may involve the formation of a loop, the reversion

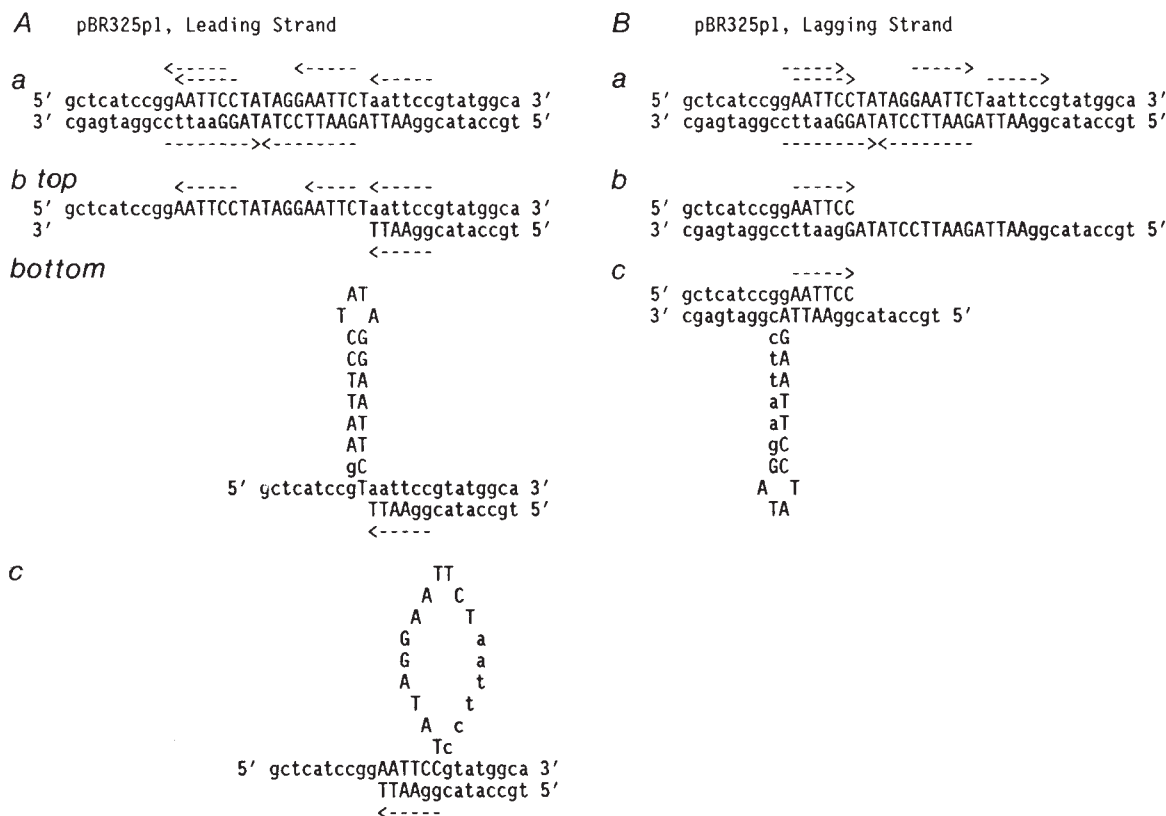


FIG. 3 Potential structural intermediates involved in deletion of a palindromic asymmetric deletion insert. A, Potential DNA secondary structure in the leading strand of pBR325p1 containing an 'asymmetric palindromic insert'. a, Nucleotide sequence of the 18-bp p1 insert and flanking DNA. b, The direct repeats are represented by arrows above the sequence and the inverted repeat is indicated by facing arrows below. b, top, DNA replication of the first copy of the direct repeat 5'AATTCC3' in single stranded DNA. b, bottom, Replication in DNA in which the inverted repeat has formed a hairpin arm. Such a structure may temporarily stop the replication fork¹⁵. Note that the second copy of the direct repeat is now base-paired within the hairpin arm. c, The template strand slips to allow mispairing between the second (left most) copy of the direct repeat on the template strand and the first copy of the direct repeat on the progeny strand. This structure may form from either intermediate shown in b (top and bottom). This misalignment event requires formation of a loop of DNA that can not form a hairpin arm. B, Potential DNA secondary structure in the lagging strand of pBR325p1.

a, Orientation of direct and inverted repeats. The arrows indicating the direct repeats are shown in the direction of replication. b, Replication of the first copy of the direct repeat. If the inverted repeat existed as a hairpin before replication, the replication fork would necessarily begin replication through the hairpin arm to copy the first direct repeat. c, Misalignment of the first copy of the direct repeat in the progeny strand with the second (right-most) copy of the direct repeat in the template strand. In this case, because of the location of the inverted repeat, the misaligned intermediate contains a perfect hairpin arm in the template strand which may stabilize the misalignment. Hairpins in cruciform structures represent Holliday recombinational intermediates in which, under the proper ionic conditions, there are no unpaired bases¹⁶. Three-way junctions have been shown to exist as stable structures and probably contain unstacked helices in a 'Y' structure¹⁷. The conformation of the DNA may well be influenced by the interaction with DNA polymerase. The structure as drawn would contain a G:A mismatch at the three-way junction.

TABLE 1 The frequency of Cm^r reversion mutations

pBR325, natural orientation			pBR523, CAT gene reversed	
	Plasmid	Reversion frequency	Plasmid	Reversion frequency
p1	pBR325p1	$44 \times 10^{-9} \pm 20 \times 10^{-9}$	pBR523p1	$2.0 \times 10^{-9} \pm 0.9 \times 10^{-9}$
n1	pBR325n1	$3.7 \times 10^{-9} \pm 0.9 \times 10^{-9}$	pBR523n1	$2.7 \times 10^{-9} \pm 0.9 \times 10^{-9}$

To determine reversion frequencies, 25 ml of *E. coli* DH5 containing the plasmids were grown to a density of 5×10^8 cells ml^{-1} in K medium containing $10 \mu\text{g ml}^{-1}$ tetracycline from 1 ml of an overnight culture. Reversion frequencies are the number of Cm^r cells in the population, which is determined by plating a known number of cells on Luria plates containing tetracycline and chloramphenicol ($25 \mu\text{g ml}^{-1}$ each). Total viable cells were determined by plating on medium containing tetracycline. The frequency represents the average of 6–8 to individual experiments in which the variation between experiments was typically within a factor of two. Analysis of plasmid copy number showed no detectable difference for any of these plasmids. The structure of the revertants was determined by digestion with *EcoRI* and analysis of the 129-bp *AluI* fragment, which contained the *EcoRI* site, to establish the size of the deletion to within about 2 bp. The DNA sequence of the region was determined for several revertants from each deletion insert. In >95% of the revertants examined for pBR325p1 (and n2 and p4 inserts) complete deletion of the insert and one copy of the direct repeat occurred. For the n1 plasmids and pBR523p1 the revertants consisted of complete deletions in addition to some revertants which have not completely deleted the insert. We examined 12–24 independent revertants for each plasmid.

frequency was about 10–20-fold lower, or $2\text{--}3.7 \times 10^{-9}$, consistent with the interpretation that misalignment stabilized by a hairpin has a higher reversion frequency than misalignment in which a loop is formed. Analysis of other inserts provides additional evidence in support of this interpretation (T.Q.T. and R.R.S., manuscript in preparation). Based on these results, if deletion between direct repeats occurred preferentially on the lagging strand, pBR325p1 should have a higher reversion frequency than pBR523p1. The results suggest that misalignment leading to deletion may occur preferentially in the lagging strand and its template.

Although our results are consistent with differential mutagenesis in the leading and lagging strands, we cannot exclude the possibility that differences in reversion frequencies are due to differences in the effects of orientation of these sequences on homologous or illegitimate recombination, or aberrant repair of spontaneous damage. As reversions frequencies were measured in *recA*[−] strains homologous recombination should be minimal and there seems no reason to expect different rates of recombination in the insert when the whole gene is reversed. In addition, if such differences did exist, they might be expected to be seen with all deletion inserts. The correlation of deletion frequencies with potential for the formation and stability of possible replication intermediates therefore strongly supports the hypothesis that the observed deletion events are the result of replication errors (T.Q.T. and R.R.S., manuscript in preparation).

Eukaryotic cells repair DNA damage in active genes more efficiently than in inactive genes¹⁰. In addition, a preference for repair of the transcribed strand has been demonstrated in both *E. coli*¹¹ and eukaryotic cells¹². As asymmetric DNA replication complexes are involved in eukaryotes¹³ and prokaryotes⁶ a difference in the rate of spontaneous mutagenesis in different strands might not be unexpected. An initial investigation did not reveal a large difference in the fidelity of misincorporation during DNA replication in an *in vitro* HeLa cells system¹⁴. Our results present the first evidence for differential mutagenesis in the leading or lagging strand of *E. coli*. □

Received 12 March; accepted 10 June 1991.

1. Glickman, B. W. & Ripley, L. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4046–4050 (1984).

- Ripley, L. S. & Glickman, B. W. in *Cellular Responses to DNA Damage*, UCLA Symposia on Molecular and Cellular Biology, Vol 10 (ed. N. R. Cozzarelli) 521–540 (Liss, inc. New York, N.Y. 1983).
- Drake, J. W., Glickman, B. W. & Ripley, L. S. *Am. Scient.* **71**, 621–630 (1983).
- Schaaper, R. M., Danforth, B. N. & Glickman, B. W. *J. molec. Biol.* **189**, 273–284 (1986).
- Ripley, L. S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4128–4132 (1982).
- Streisinger, G. et al. *Cold Spring Harb. Symp. quant. Biol.* **31**, 77–84 (1966).
- Albertini, A. M., Hofer, M., Calos, M. P. & Miller, J. H. *Cell* **29**, 319–328 (1982).
- McHenry, C. S. A. *Rev. Biochem.* **57**, 519–550 (1988).
- Scott, J. R. *Microbiol. Rev.* **48**, 1–23 (1984).
- Bohr, V. A., Smith, C. A., Okumoto, D. A. & Hanawalt, P. C. *Cell* **40**, 359–369 (1985).
- Mellon, I. & Hanawalt, P. C. *Nature* **342**, 95–98 (1989).
- Mellon, I., Spivak, G. & Hanawalt, P. C. *Cell* **51**, 241–249 (1988).
- Tsurimoto, T., Melendy, T. & Stillman, B. *Nature* **346**, 534–539 (1990).
- Roberts, J. D., Thomas, D. C. & Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3465–3469 (1991).
- Weaver, D. T. & DePamphilis, M. L. *J. molec. Biol.* **180**, 961–986 (1984).
- Duckett, D. R., Murchie, A. I. H. & Lilley, D. M. J. *EMBO J.* **9**, 583–590 (1990).
- Duckett, D. R. & Lilley, D. M. J. *EMBO J.* **9**, 1659–1664 (1990).

ACKNOWLEDGEMENTS. This work was supported by grants from the American Cancer Society and the National Institutes of Environmental Health Sciences. We thank K. Allen and E. Schwab for assistance.

ERRATUM

Homozygous prion protein genotype predisposes to sporadic Creutzfeldt–Jakob disease

Mark S. Palmer, Aidan J. Dryden, J. Trevor Hughes & John Collinge

Nature **352**, 340–342 (1991)

IN this letter in the 25 July issue, the sentence beginning in line 1 on page 341 should read: “Of the 22 CJD cases, 16 were homozygous for Met 129, 5 homozygous for Val 129 and only one was a heterozygote, while of the suspected CJD cases, 13 were Met-129 homozygotes, 4 were heterozygotes and 6 were Val-129 homozygotes.” The published version incorrectly referred to 11 Met-129 homozygotes.

CORRECTION

Novel myosin heavy chain encoded by murine *dilute* coat colour locus

John A. Mercer, Peter K. Seperack, Marjorie C. Strobel, Neal G. Copeland & Nancy A. Jenkins

Nature **349**, 709–713 (1991)

WE have discovered frameshift errors made in the assembly of our published cDNA sequence affecting only the region of nucleotides 1138–1180, corresponding to amino acid residues 367–379. The correct sequence is available on the EMBL database under accession number X57377, release 28. These errors do not affect the conclusions of the paper in any way. We thank D. Cheney, E. Espreafico, R. Larson, and M. Mooseker for sharing unpublished sequence data.

Solid-phase gene assembly

Kenneth L. Beattie and Richard F. Fowler

New technologies for multiple chemical synthesis of oligonucleotides and stepwise hybridization on a solid-phase support enable the rapid and cost-effective preparation of long duplex DNA regions. Will these new technologies usher in a new era in protein engineering?

TEN years ago, few laboratories had access to synthetic DNA as a research tool. Because of entrepreneurial scientific and business activities in the 1980s synthetic DNA became widely accessible to the research community, through automated instrumentation for chemical synthesis of DNA and the low-cost commercial supply of oligonucleotides. Now, virtually every molecular biology laboratory uses synthetic DNA as an essential reagent. Fundamental research capabilities that are used widely because of the availability at low cost of synthetic DNA include dideoxy DNA sequencing¹, site-directed mutagenesis² and the polymerase chain reaction (PCR)³. In addition, the availability of synthetic DNA has spawned basic research, clinical applications and commercial activities in DNA probe diagnostics^{4,5} and DNA therapeutics^{6,7}.

Gene synthesis was the first research application of synthetic DNA. This was achieved in Khorana's laboratory⁸ using solution-phase phosphotriester chemistry long before the advent of automated solid-phase DNA synthesizers. Potentially, gene synthesis is a powerful research capability. *De novo* gene design, starting with the nucleotide sequence of 'naturally occurring' genes, can incorporate advantageous features, including (1) optimization of codon usage for improved gene expression in a heterologous host; (2) positioning of single-cutting restriction sites at convenient intervals across a gene to facilitate subsequent manipulation of the gene by recombinant DNA technology; (3) insertion or substitution of a coding region from another gene (to create chimaeric or fusion proteins); and (4) introduction of a section containing one or more desired mutations.

Despite the potential usefulness of gene synthesis in protein engineering, as yet the technique has not become a widely used tool in molecular biology research for the same reason that kept synthetic DNA from being used widely until recently: inaccessibility of the technology due to the high cost of materials and laborious procedures. The cost of synthetic DNA needed for the assembly of an average size gene currently precludes the routine use of gene synthesis in most laboratories. In addition, the gene assembly process, which typically

involves annealing, ligation and cloning of overlapping sets of complementary oligonucleotides, can often become a significant research project, requiring months of work. Amplification of the assembled gene by PCR may facilitate recovery of the intact gene⁹ and reduce the assembly time. However, a problem that commonly accompanies gene assembly is mutation, necessitating further corrective manipulations of the gene.

Technological advances

Two recent technological developments should minimize the technical and cost limitations of gene synthesis and make gene assembly a more accessible research tool. The first is a process for the rapid and cost-effective simultaneous synthesis of numerous oligonucleotides¹⁰, using a segmented synthesis process that improves upon a previous filter paper disc method^{11,12}. 'Synthesis wafers' consist of Teflon cylinders containing nucleoside-derivatized porous glass, retained within the wafer by porous Teflon mesh. As illustrated in figure 1, wafers are stacked into columns to allow simultaneous addition of A, G, C or T to all DNA chains within a given stack. A different activated nucleoside phosphoramidite¹³ is intro-

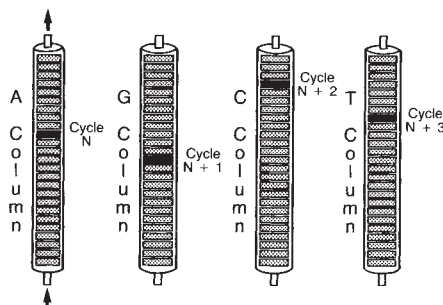


FIG. 1 Simultaneous chemical synthesis of numerous oligonucleotides by segmented synthesis within porous Teflon wafers. Nucleoside-derivatized controlled pore glass is placed within individual wafers, which are then stacked into columns. Simultaneous addition of A, G, C or T to DNA chains retained within the wafers is accomplished by the sequential flow of reagents through the columns by the phosphoramidite method. Following each reaction cycle, the wafers are sorted into different columns to provide for synthesis of a different nucleotide sequence within each wafer. The sequential positioning of a wafer through four cycles is depicted, resulting in the addition of AGCT to the sequence.

duced into each column. The reaction cycle is completed by sequentially pumping additional reagents through all the columns to add another base to each sequence, and then the wafers are sorted into different columns to provide for synthesis of a different sequence within each wafer. Thus, in order to add the sequence AGCT to a growing DNA chain, the wafer would be positioned in the A column during cycle N, in the G column during cycle N+1, in the C column during cycle N+2 and in the T column during cycle N+3. Full automation of this process is under development. A prototype instrument, operating at 0.1–1 μ mol scale, can produce 50–100 oligonucleotides within 3–6 hours, at a 30–50 per cent reduction in reagent scale consumption as compared to standard DNA synthesis instruments. Using this technology, it is possible to produce all of the oligonucleotides needed to assemble an average size (1,000 base pairs) gene within a single day. The reagent cost should be reduced significantly by implementing a 10–20 nmol scale synthesis.

A solid-phase gene assembly process^{10,14} is the second technological advance that promises to enable economical, rapid and efficient gene synthesis. The concept, illustrated in figure 2, is analogous to the solid-phase synthesis of DNA, except that instead of conducting stepwise addition of individual nucleotide residues, stepwise addition of oligonucleotides is carried out to build up a long duplex DNA region. Gene assembly is initiated with a support-bound oligonucleotide, to which successive oligonucleotides (5'-phosphorylated and added at molar excess) are annealed. Controlled, stepwise annealing and washing (requiring a few minutes per oligonucleotide) results in the rapid and accurate assembly of long duplex DNA. The completed duplex DNA can be ligated on the support, and design of the gene accommodates its release from the support by digestion with a restriction enzyme. The duplex segment can be directly cloned, sequenced and expressed. Although thus far, the solid-phase approach has been used to assemble a relatively small number of genes, it is clear that solid-phase gene assembly is a reliable and rapid method for construction of duplex DNA several

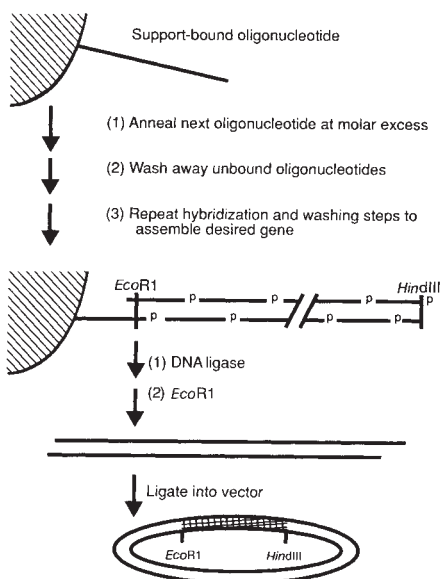


FIG. 2 Schematic illustration of solid-phase gene assembly. The desired gene is designed to be assembled from a set of overlapping complementary oligonucleotides. Assembly is initiated with an oligonucleotide bound at one end to a solid-phase support. 5'-phosphorylated oligonucleotides are added sequentially (at molar excess) to the support-bound oligonucleotide. Unbound DNA is washed away before each successive annealing reaction. The completed assembly is treated with DNA ligase to seal the nicks, then the gene is released from the support by cleavage at a unique restriction site contained within the support-bound oligonucleotide. The released DNA is ligated into a suitable vector for sequencing and expression.

hundred base pairs in length. Sequencing of cloned products has indicated an exceptionally low mutation frequency, an apparent advantage over solution-phase assembly procedures which commonly yield mutations, possibly due to heterogeneous DNA structures formed when numerous oligonucleotides are annealed together.

The support material that is used at present for solid-phase gene assembly consists of paramagnetic latex microspheres coated with streptavidin (Dynal Corporation, Great Neck, New York) to which 5'-biotinylated oligonucleotides bind tightly. Stepwise annealing is carried out at 45°C in buffer containing 1 M NaCl, using enzymatically phosphorylated oligonucleotides 30–40 residues in length. The avidin-biotin linkage is useful for the assembly of duplex DNA 500–800 bp in length, but long duplex DNAs (> 1,000 bp) end-labelled with biotin are not bound tightly to streptavidin-coated beads. This problem can be circumvented, however, by initiating the assembly from both ends of the gene and joining the gene fragments after release from the support. Oligonucleotides covalently linked to other supports (for example, glass) are currently under

evaluation for initiation of longer gene assemblies. Although 15–20-bp overlaps provide reliable hybridization in solid-phase gene assembly, the use of shorter oligonucleotides in the assembly is being evaluated. If solid-phase gene assembly could be accomplished with hexamers (possibly requiring the assistance of DNA ligase to stabilise each addition), then a library of all 4,096 6-mers could be used for 'off-the-shelf' assembly of any possible gene.

Automation of solid-phase gene assembly could readily be accomplished, and a gene assembly instrument could be supplied with versatile computer software for gene design, together with devices for simultaneous processing and purification of numerous oligonucleotides. Yet, despite the prospect of extensive automation, gene synthesis may never be an absolutely reliable, fool-proof process, as the necessity of cloning, sequencing and expressing synthetic genes adds a potentially unpredictable research component to the process.

Visions of the future

DNA sequence information that will come from the Human Genome Project will define numerous targets for gene synthesis and protein engineering. Solid-phase gene assembly can play a major role in future applications of genome sequence information in medicine and agriculture, provided that the technology, in the form of instrumentation, is made widely accessible to the research community. By analogy to the numerous powerful research capabilities that arose from the widespread availability of synthetic DNA, it is anticipated that affordable and rapid gene assembly will likewise encourage imaginative new applications of the technology, particularly in the fields of protein engineering and biotechnology. Developments that are perceived to be within our technological grasp include: (1) synthesis of oligonucleotide libraries for protein engineering, enabling the assembly of essentially any desired gene; (2) creation of libraries of duplex DNA segments encoding protein molecules¹⁵, structural motifs, genetic signals, and other sequence elements that could serve as a resource for the modular assembly of a very large variety of novel genes; (3) parallel assembly of an array of duplex DNAs on a two-dimensional surface, enabling the simultaneous assembly of numerous variants of a given gene, or the assembly of numerous duplex regions that subsequently could be joined to form an extensive duplex DNA; and (4) assembly of synthetic viral-like genomes that could serve as pharmaceutical agents. □

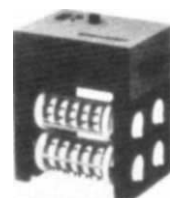
Kenneth L. Beattie and Richard F. Fowler are at the Center for Biotechnology, Baylor College of Medicine, 4000 Research Forest

Drive, The Woodlands, Texas 77381, USA. K.L.B. is also employed by the Houston Advanced Research Center, 4800 Research Forest Drive, The Woodlands, Texas 77381, USA. Commercial development of the multiple DNA synthesis and solid-phase gene assembly technologies is taking place at Genosys Biotechnologies, Inc., 8701 New Trails Drive, The Woodlands, Texas 77381, USA. For more information, fill in reader service number 100.

1. Sanger, F., Niklen, S. & Coulson, A.R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
2. Zoller, M.J. & Smith, M. *DNA* **3**, 479–488 (1982).
3. Mullis, K.B. & Faloona, F.A. *Meth. Enzym.* **155**, 335–350 (1987).
4. Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L. & Wallace, B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 278–282 (1983).
5. Saiki, R.K., Walsh, P.S., Levenson, C.H. & Erlich, H.A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6230–6234 (1989).
6. Uhlmann, E. & Peyman, A. *Chem. Rev.* **90**, 543–584 (1990).
7. Riordan, M.L. & Martin, J.C. *Nature* **350**, 442–443 (1991).
8. Khorana, H.G. *Science* **203**, 614–625 (1979).
9. Jayaraman, K., Fingar, S.A., Shah, J. & Fyles, J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4084–4088 (1991).
10. Beattie, K.L. et al. *Biotechnol. appl. Biochem.* **10**, 510–521 (1988).
11. Frank, R., Heikens, W., Heisterberg-Moutsis, G. & Blocker, H. *Nucleic Acids Res.* **11**, 4365–4377 (1983).
12. Matthes, H.W.D., Zenke, W.M., Grundstrom, T., Staub, A., Wintzerth, M. & Chambon, P. *EMBO J.* **3**, 801–805 (1984).
13. McBride, L.J. & Caruthers, M.H. *Tetrahedron Lett.* **24**, 245–248 (1983).
14. Hostomsky, Z., Smrt, J., Arnold, L., Tocik, Z. & Paces, V. *Nucleic Acids Res.* **15**, 4849–4856 (1987).
15. Baron, M., Norman, D.G. & Campbell, I.D. *Trends biochem. Sci.* **16**, 13–17 (1991).

ADVERTISEMENTS

DIABORM EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

DIABORM
Gentle
P.O. Box 550/26
P-8900 Munich 65.
FRG.

Reader Service No.49

Custom Peptides

\$30 a residue, 10 µ mole scale
\$80 a residue, 100 µ mole scale

Research Genetics

1-800-533-4363

In the United Kingdom 0-800-89-1393

In Canada 1-800-533-4363

FAX 205-536-9016

Reader Service No.32

ADVERTISEMENTS

DNA

Around the World

\$4.50 a base
(no set-up)

Includes:

- 500-1500 ugs
- 15-45 O.D.S.
- 24-72 hour delivery
- Amplification ready
- Gel picture
- Quality guaranteed products

We are a full service and design DNA lab.

1-800-227-0627
Fax 1-214-420-0442
BIO-SYNTHESIS

Reader Service No.18

TIB MOLBIOL BERLIN
CUSTOM DNA SYNTHESIS

5,- DM per base

DM 95,- per oligo

FAX (0049) 30 6261960

Reader Service No.79

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.17

Molecular miscellany

A recombinant T4 gene 32 protein, antisense phosphorothioate oligos and a second generation capillary electrophoresis system with a thermostating capability — new product ideas from the molecular biology marketplace.

DNA sequencing tasks, large or small, are now being accepted by United States Biochemical as part of a newly launched **custom DNA sequencing service** (*Reader Service No. 101*). Long-term project commitments are welcomed. The sequencing service includes complete data analysis, as required. Full restriction maps can be generated from the data. USB will determine nucleotide and codon frequency, and perform translation and open reading frame analysis. GENBANK can also be screened. The analysis service is offered at no extra cost. Researchers will receive the original autoradiograms along with the results on disk (DOS or Macintosh format). USB will also return any subclones/deletion mutants or primers used to determine the sequence. USB says all disclosures made to the company are strictly confidential and can be formalized with a non-disclosure agreement. In addition, USB says the prices it charges for sequencing DNA samples are competitive with in-house sequencing and quotes will be given on an individual project basis.

The removal of single-stranded DNA from miniprep isolations is often a necessary procedure for research involving recombinant DNA and sequencing. Schleicher and Schuell has introduced the Centrex **disposable centrifugal filtration device**, which is designed to provide a quick and easy method of separating single-stranded from double-stranded DNA (*Reader Service No. 102*). Exploiting the affinity of nitrocellulose for single-stranded DNA, S & S's NC nitrocellulose binds preferentially to single-stranded DNA, allowing it to be removed from a double-stranded DNA sample. When using S & S's Centrex disposable centrifugal filtration device with a 0.2- μ m nitrocellulose filter, con-

taminating single-stranded DNA is retained on the filter allowing the double-stranded DNA to pass through for collection in a receiving tube.

PCR protocols

The Genesphere **mini clean room** from Safetech provides all-round visibility for the operator and incorporates an ultraviolet light irradiation facility (*Reader Service No. 103*). As such, it is intended



Mini clean room with UV irradiation.

for use in applications involving DNA amplification using PCR, which is acutely sensitive to cross-contamination. The special 'purge' mode, bathes the inside of the Genesphere in ultraviolet light; this function is designed to destroy contaminating DNA and eliminate the risk of amplifying contaminated DNA with the DNA under study. The device can be preprogrammed for overnight purging. Once purged, the Genesphere provides a flow of class 100 clean air and a contamination-free environment within the sphere, Safetech says. The Genesphere's HEPA filter can be changed in minutes. Accessories include a mobile telescopic trolley and a hi-hat extension for use with long pipettes.

Ambion announces that **recombinant T4 gene 32 protein** (GP 32) is now available (*Reader Service No. 104*). GP 32, a single-stranded DNA binding protein encoded by bacteriophage T4, has several practical applications in molecular biology. Recent reports in the scientific literature indicate that GP 32 can improve both the yield and specificity of PCR reactions, Ambion says. GP 32



S&S's spin filtration device.

binds single-stranded DNA in a cooperative manner and acts as a helix destabilising protein. According to Ambion, in experiments carried out by the company, GP 32 was found to improve the results in about half of the PCR reactions in which it was tested, including ones in which it was difficult to obtain satisfactory results by other means. Ambion says that GP 32 frequently yields the desired result in a single experiment, which otherwise could be obtained only by systematically testing variables such as magnesium, nucleotide, primer and template DNA concentrations. In addition, GP 32 has been reported to reduce background bands in DNA sequencing reactions and to help drive large-scale restriction enzyme digests to completion. GP 32 is supplied as a highly purified preparation, free of detectable nucleases, in 100- and 500- μ g sizes.

Multi Technology has introduced a new generation of **filter tips** called Multi Guard, which incorporates a specially designed hydrophobic barrier to prevent airborne biohazards, toxic chemicals and radioactivity from contaminating pipetors during liquid handling procedures (Reader Service No. 105). MTI claims that use of Multi Guard tips will eliminate false signals after PCR amplification due to cross-contamination. There are six versions of pipette tip to choose from.

Assorted reagents

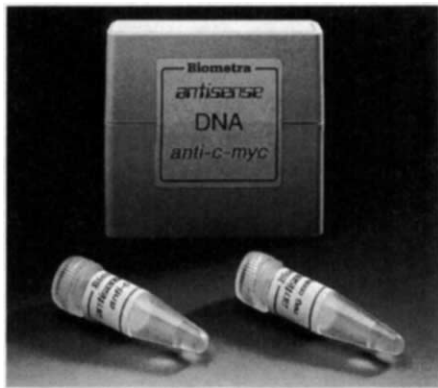
Biometra has extended its range of **antisense phosphorothioate oligonucleotides** (PTOs) for the specific inhibition of translation *in vivo* (Reader Service No. 106). These modified PTOs are highly resistant to nuclease attack and therefore eliminate many of the problems associated with using standard oligonucleotides, Biometra says. In addition, they exhibit similar, or even better, hybridization characteristics in cellular conditions than their standard oligonucleotide equivalents. Biometra's antisense PTOs are supplied in a ready-to-use form that can be added directly to the cell culture medium. Cellular uptake occurs directly

from the medium; no treatment of cells or medium is required to enhance uptake, Biometra says. As a complement to the antisense PTOs introduced earlier, which included anti-*c-myc*, anti-*c-fos*, anti-PDGF and anti-TGF, Biometra now offers new antisense PTOs against proto-oncogenes (*c-mos*, *c-myc*, *c-src*, *c-yes* and *c-ets*) and against growth factors (IGF, CSF, G-CSF, GM-CSF and their corresponding receptors). In addition, several antisense PTOs against interleukins and intestinal polypeptides (VIP, ICaBP, FABP and TNF) are now available.

Perform high fidelity DNA synthesis with an error rate that is 12-fold less than with *Taq* DNA polymerase. That is the claim made by Stratagene when using the company's *Pfu* DNA polymerase, a multifunctional, **thermostable enzyme** that possesses both a 5' to 3' DNA polymerase activity and a 3' to 5' exonuclease proofreading activity (Reader Service No. 107). During DNA synthesis, the 3' to 5' exonuclease activity of the *Pfu* DNA polymerase will excise mismatched nucleotides from the 3' terminus of the primer: template complexes and permits the polymerase activity to incorporate the correct nucleotide complementary to the template strand. The 3' to 5' exonuclease activity does not degrade single-stranded DNA, Stratagene says. In addition, the activity versus temperature profile of *Pfu* DNA polymerase is comparable to that of *Taq* DNA polymerase.

Gibco-BRL has introduced two new standards for the sizing of large DNA (Reader Service No. 108). MegaBase II DNA Standard can be used for sizing 48.5 kb to 1.3 Mb DNA fragments. For sizing DNA fragments in the 3.5–5.7-Mb range, the company offers MegaBase III DNA Standard. MegaBase II DNA Standard is a set of lambda concatamers in blue agarose beads. As the precise size of the lambda monomer unit is well established, accurate size determinations can be made with this standard, the company says. The second DNA standard, MegaBase III, is a set of three *Schizosaccharomyces pombe* chromosomes (approximately 3.5, 4.7 and 5.7 Mb in size) encapsulated in green agarose beads. With both products, the beads are easy to see during gel loading. The agarose bead configuration used with these products is designed to be more convenient than agarose plugs. After electrophoresis, the bands are visualized by staining with ethidium bromide.

Fisons has eliminated the need to handle corrosive phenol solutions with the introduction of **pre-equilibrated phenol solution** for molecular biology appli-



Antisense phosphorothioate oligos.

ADVERTISEMENT

DNA FINGERPRINTING

Fresenius AG, Bad Homburg is offering the following oligonucleotides **exclusively**, which were evaluated and **patented** by Dr. J.T. Epplen and colleagues, Max-Planck-Institut für Psychiatrie, Martinsried, F.R.G.:

(GTG)₅ · (GATA)₄ · (GACA)₄ (GGAT)₄ · (TCC)₃

These probes are available in unlabeled (10 pMol/aliquot) or Digoxigenin labeled (140 pMol/aliquot) form; lyophilized.

Fresenius

Diagnostic Division

6370 Oberursel, F.R.G. · Telephone: +61 71/60-2443
Telefax: +61 71/60-21 35 · Postbox 1809

Reader Service No.78

Synthetic Oligonucleotides

Custom made probes and primers, synthesised, purified and shipped in 3 days

£3.50 per base

(All inclusive)

0.2 μ M scale (approx. 500 μ g)

Modified bases, blotin or digoxigenin labelling, data base analysis and full design facilities

Molecular Medicine Unit,
King's College School of Medicine & Dentistry,
Denmark Hill, London SE5 8RX

Tel. XX-44-(0)71 - 326 3126

FAX XX-44-(0)71 - 733 3877

Reader Service No.56

TRITIATED & ULTRA PURE PEPTIDES

An increasingly recognized problem is the purity of peptides used in biological studies. We offer synthesis of peptides including chemical synthesis of tritiated peptides of a very high purity.

SERTEK

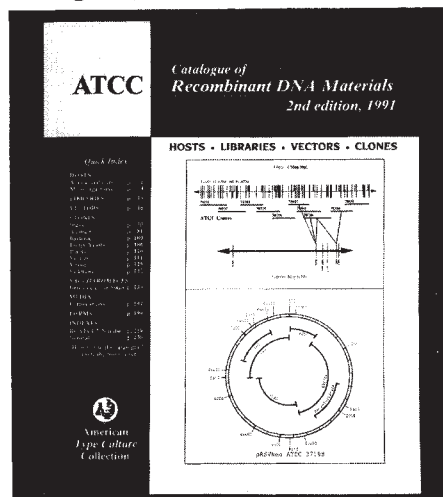
Haraldsgade 68, DK-2100 Denmark
Tel: +45-3120 2554; Fax +45-3181 7833

Reader Service No.15

cations (*Reader Service No. 109*). The preparation involved in the production of equilibrated phenol includes the addition of 8-hydroxyquinoline to a saturated phenol solution (0.1% w/v). This is then extracted, initially with a 1 M Tris buffer solution, followed by a 0.1 M Tris buffer solution until the pH of the aqueous phase is greater than 7.6. The solution is sold in 100-ml aliquots costing £5.50 (UK). It should be stored at -20 °C protected from light.

New product listings

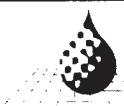
Over 1,350 different recombinant DNA items are listed in the second edition of the American Type Culture Collection's (ATCC's) **recombinant DNA materials catalogue** (*Reader Service No. 110*).



One-stop shopping with ATCC.

This 240-page catalogue lists cloning vectors; hosts for transformation and transfection (animal cell lines and microorganisms); clones from animal, bacterial, fungal, yeast, plant, viral and viroid genomes; libraries; and genomic clone maps of *Saccharomyces cerevisiae*. A companion catalogue, entitled the ATCC/NIH repository of human and mouse DNA probes and libraries catalogue (4th edition), lists the probes, clones and chromosome-specific libraries from the human genome that are available through ATCC.

ADVERTISEMENT



Severn Biotech Ltd.

TM

Custom DNA Synthesis
Probes • Primers • Oligo's

**For a fast efficient
Cost effective service**

BUY SEVERN BIOTECH

Tel: 0562 825286 Fax: 0562 825284

Unit 23, Hoobrook Enterprise Centre,
Worcester Road, Kidderminster DY10 1HY

Reader Service No.20

Enzyme Express, the result of a joint venture between Scientific Products and Pharmacia, is a new distribution arrangement for the **rapid delivery of high-quality restriction and modifying enzymes** (*Reader Service No. 111*). The extensive range of Pharmacia's enzymes, available exclusively through Scientific Products, is outlined in the new 56-page catalogue of restriction and modifying enzymes for molecular biology. Enzyme Express offers delivery within 24 hours anywhere in the United States through its nationwide network of 26 distribution centres. The catalogue features 58 high-quality restriction enzymes, some with rare recognition sequences, which are useful for generating large fragments from genomic DNAs, the company says. Pharmacia's line-up also includes modifying enzymes that can be used in applications such as sequencing, cloning, labelling, mutagenesis, transcription, translation and synthesis experiments.

Laboratory hardware

Sample evaporation control and a thermostating capability are features of the Model 270A-HT **high-throughput capillary electrophoresis system**, a second generation system from Applied Biosystems (*Reader Service No. 112*). The Model 270A-HT provides 50 sample positions and eight buffer reservoirs. Effective evaporation control has been shown to be a critical factor in run-to-run reproducibility and sample quantification. Applied Biosystems says the efficiency of the closure design has been tested extensively under rigorous CE separation conditions, which have included the addition of many volatile organic additives during extended overnight operation. A reproducibility of run-to-run migration times of less than one per cent is routine, says the company. The ability to cool samples that are awaiting analysis to 4 °C or less helps to ensure the integrity of thermolabile compounds; this is an important consideration as samples can sit in an autosampler for as many as 12 to 18 hours before analysis. In addition to the cooling option, samples can also be heated. This opens up a wealth of new applications for CE, including its use to monitor the time-course of protein digests or to investigate the kinetics of enzyme catalysis in an automated mode.

The Hitachi HDG-1717BL **back-lighted desktop tablet digitizer** is designed for reading DNA sequencing films, autoradiography-based X-ray films or other translucent films/medias (*Reader Service No. 113*). With a digitizing surface of 17×17 inches, the device features a built-in light box, a stylus pen for data input and variable (60–100 per cent) light adjustment. The \$2,800 (US) HDG-



Hitachi's back-lit digitizer speeds DNA base input.

1717BL digitizer uses electromagnetic induction to determine the exact location of input points. The input points are indicated by touching the stylus pen to the digitizing surface. Hitachi claims that the electromagnetic induction method of point detection is not subject to the same sound-wave disturbances as the ultrasonic triangulation method used by other digitizers. The digitizer is designed for use in conjunction with the company's software packages for DNA and protein sequence analysis, such as the HIBIO DNASIS and HIBIO PROSIS. However, it can also be used with other IBM- and Macintosh-compatible software packages.

Stuart Scientific announces the availability of its **GENE-TECH thermal cycler**, which made its debut at Achema, Germany in June (*Reader Service No. 114*). This fully self-contained unit has built-in peltier thermo-electric modules, which provide temperature control from 4 °C to 99 °C, without the need for an ancillary thermocirculator, Stuart says. The GENE-TECH can accommodate 40×0.5-ml microcentrifuge tubes or one microtitre plate, as required. Temperature can be ramped up, down or held according to



Stuart Scientific's GENE-TECH thermal cycler takes tubes or plates.

the user's choice of programmes. There are 30 programmes with 29 segments each; all are fully linkable. For operational simplicity, 10 programmes have been preset. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated. The polymerase chain reaction (PCR) process is covered by US patents issued to Cetus Corporation.